

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number  
**WO 02/102984 A2**

- (51) International Patent Classification<sup>7</sup>: C12N (74) Agents: **BROOK, David, E.** et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).
- (21) International Application Number: PCT/US02/19051
- (22) International Filing Date: 14 June 2002 (14.06.2002) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/298,173 14 June 2001 (14.06.2001) US  
60/311,686 10 August 2001 (10.08.2001) US  
60/316,995 4 September 2001 (04.09.2001) US
- (71) Applicant (*for all designated States except US*): **SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH** [US/US]; 1275 York Avenue, New York, NY 10021 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **RICHON, Victoria** [US/US]; 160 Theodore Fremd Street, #A11, Rye, NY 10580 (US). **ZHOU, Xianbo** [CN/US]; 43 Bradley Street, Dobbs Ferry, NY 10522 (US). **RIFKIND, Richard, A.** [US/US]; 425 East 58th Street, #48A, New York, NY 10022 (US). **MARKS, Paul, A.** [US/US]; 7 Rossiter Road, Washington, CT 06793 (US).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:**  
— *without international search report and to be republished upon receipt of that report*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 02/102984 A2

(54) Title: HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

(57) Abstract: The present invention features substantially pure HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), an HDRP( $\Delta$ NLS) polypeptides, and isolated nucleic acid molecules encoding those polypeptides. The present invention also features vectors containing HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) nucleic acid sequences, and cells containing those vectors.

## HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

## RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No.

- 5 60/298,173 filed on June 14, 2001, U.S. Provisional Application No. 60/311,686  
filed on August 10, 2001, and U.S. Provisional Application No. 60/316,995, filed on  
September 4, 2001. The entire teachings of the above applications are incorporated  
herein by reference.

## 10 GOVERNMENT SUPPORT

The invention was supported, in whole or in part, by grant CA-0974823 from  
the National Cancer Institute. The Government has certain rights in the invention.

## BACKGROUND OF THE INVENTION

- 15 The N-terminal tails of core histones are covalently modified by post-  
translational modifications, including acetylation and phosphorylation. Evidence  
suggests that these covalent modifications play important roles in several biological  
activities involving chromatin, *e.g.*, transcription and replication. Histone  
deacetylases (HDACs) catalyze the removal of the acetyl group from the lysine  
20 residues in the N-terminal tails of nucleosomal core histones resulting in a more  
compact chromatin structure, a configuration that is generally associated with  
repression of transcription.

- Five proteins and/or open reading frames in yeast (RPD3, HDA1, HOS1,  
HOS2 and HOS3) that share significant homology in the catalytic domain have been  
25 identified as HDACs based upon their sequence homology to human HDAC1. To  
date, eight HDACs have been identified in mammalian cells, and classified into two  
classes based on their structure and similarity to yeast RPD3 or HDA1 proteins.  
Recently, Sir2 family proteins that are structurally unrelated to the five proteins  
aforementioned have been identified as NAD-dependent HDACs. Class I HDACs  
30 are the yeast RPD3 homologs HDAC1, 2, 3, and 8, and are composed primarily of a  
catalytic domain. Class II HDACs are the yeast HDA1 homologs HDAC4, 5, 6; and

7. HDAC4, 5, and 7 contain a long non-catalytic N-terminal end and a C-terminal HDAC catalytic domain while HDAC6 has two HDAC catalytic domains.

It has also been determined that histone deacetylases can be sensitive to small molecules, including trichostatin A (TSA), trapoxin, and butyrate. For example, the yeast RPD3 and HDA1 and mammalian HDAC1, 2, 3, 4, 5, 6, 7 and 8 are sensitive to inhibition by trichostatin A (TSA). The Sir2 family HDACs, yeast HOS3 and *Drosophila melanogaster* dHDAC6, however, appear to be relatively insensitive to TSA. A class of hybrid bipolar compounds, such as suberoylanilide hydroxamic acid (SAHA) have also been shown to inhibit histone deacetylases and induce terminal differentiation and/or apoptosis in various transformed cells. Examples of such compounds can be found in U.S. Patent Nos. 5,369,108, issued on November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, as well as U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, the entire content of all of which are hereby incorporated by reference.

The identification of the mechanisms by which histones are deacetylated, and the characterization of histone deacetylase function would be of great benefit in understanding how gene transcription is controlled, how the cell cycle is regulated, and how cells are signaled to undergo terminal differentiation and/or apoptosis. Elucidation of such mechanisms can lead to improved therapeutics for many diseases, in particular those characterized by cell proliferation or a lack of cell differentiation or apoptosis, for example, cancer.

#### SUMMARY OF THE INVENTION

25 The present invention relates to isolated or recombinant histone deacetylase polypeptides, and isolated histone deacetylase nucleic acid molecules encoding those polypeptides, as well as vectors and cells containing those isolated nucleic acid molecules.

In one aspect of the invention, the isolated or recombinant histone deacetylase polypeptide is selected from a) an isolated or recombinant polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and b) a polypeptide having at least 60% sequence identity with any one

of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In one embodiment, the isolated or recombinant histone deacetylase polypeptide consists of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. In another embodiment, the isolated or recombinant histone deacetylase polypeptide is mammalian; preferably, the isolated or recombinant histone deacetylase polypeptide is human.

In another aspect, the invention features an isolated nucleic acid molecule selected from a) an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; b) a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; c) an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; d) a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; e) a nucleic acid that is hybridizable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; or f) a nucleic acid molecule that is hybridizable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; and g) an isolated nucleic acid molecule that has at least 55% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof. In one embodiment, the isolated nucleic acid molecule consists of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. In another embodiment, the isolated nucleic acid molecule is mammalian; preferably, the isolated nucleic acid molecule is human.

In other aspects, the invention features a vector comprising the isolated histone deacetylase nucleic acid molecule described above, a cell comprising the vector, and a cell comprising the isolated histone deacetylase nucleic acid molecule described above.

In another aspect, the invention features a purified antibody that selectively binds a histone deacetylase polypeptide described above.



In yet another aspect, the invention features a method of identifying a compound that modulates expression of a histone deacetylase nucleic acid molecule described above. The method comprises the steps of a) contacting the nucleic acid molecule with a candidate compound under conditions suitable for expression; and  
5 b) assessing the level of expression of the nucleic acid molecule. A candidate compound that increases or decreases expression of the nucleic acid molecule relative to a control is a compound that modulates expression of the nucleic acid molecule. In one embodiment, the method is carried out in a cell or animal. In another embodiment, the method is carried out in a cell free system.

10 The invention also features a method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, for example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer and myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid  
15 metaplasia, and chronic myelogenous leukemia in an individual, comprising administering a compound identified by the above method.

In still another aspect, the invention features a method of identifying a compound that modulates the enzymatic activity of the histone deacetylase polypeptide described above. The method comprises the steps of a) contacting the  
20 polypeptide with a candidate compound under conditions suitable for enzymatic reaction; and b) assessing the activity level of the polypeptide. A candidate compound that increases or decreases the activity level of the polypeptide relative to a control is a compound that modulates the enzymatic activity of the polypeptide. In one embodiment, the method is carried out in a cell or animal. In another  
25 embodiment, the method is carried out in a cell free system.

In yet another embodiment, the polypeptide is further contacted with a substrate for the polypeptide, wherein the substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent. In one embodiment, the  
30 candidate compound is an inhibitor. In another embodiment, candidate compound is an activator.

In another aspect, the invention features a method of identifying a compound that modulates the transcriptional repression activity of the histone deacetylase polypeptide described above. The method comprises the steps of a) contacting the polypeptide with a candidate compound under conditions suitable for a  
5 transcriptional repression reaction; and b) assessing the transcriptional repression activity level of the polypeptide. A candidate compound that increases or decreases the transcriptional repression activity level of the polypeptide relative to a control is a compound that modulates the transcriptional repression activity of the polypeptide. In one embodiment, the method is carried out in a cell or animal. In another  
10 embodiment, the method is carried out in a cell free system.

In yet another embodiment, the polypeptide is further contacted with a substrate for the polypeptide, wherein the substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent. In one embodiment, the  
15 candidate compound is an inhibitor. In another embodiment, candidate compound is an activator.

In another aspect, the invention features a method of identifying a compound that modulates expression of a histone deacetylase nucleic acid molecule described above. The method comprises the steps of a) providing a nucleic acid molecule  
20 comprising a promoter region of the histone deacetylase nucleic acid molecule described above, or part of such a promoter region, operably linked to a reporter gene; b) contacting the nucleic acid molecule or with a candidate compound; and c) assessing the level of the reporter gene. A candidate compound that increases or decreases expression of the reporter gene relative to a control is a compound that  
25 modulates expression of the histone deacetylase nucleic acid molecule described above. In one embodiment, the method is carried out in a cell.

In still another aspect, the invention features a method of identifying a polypeptide that interacts with a histone deacetylase polypeptide described above in a yeast two-hybrid system. The method comprises the steps of a) providing a first  
30 nucleic acid vector comprising a nucleic acid molecule encoding a DNA binding domain and the histone deacetylase polypeptide described above; b) providing a second nucleic acid vector comprising a nucleic acid encoding a transcription

activation domain and a nucleic acid encoding a test polypeptide; c) contacting the first nucleic acid vector with the second nucleic acid vector in a yeast two-hybrid system; and d) assessing transcriptional activation in the yeast two-hybrid system. An increase in transcriptional activation relative to a control indicates that the test polypeptide is a polypeptide that interacts with the histone deacetylase polypeptide described above.

The invention also features a pharmaceutical composition comprising a histone deacetylase polypeptide described above.

In addition, the present invention features a method of diagnosing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease in a subject. The method comprises the steps of a) obtaining a sample from the subject; and b) assessing the level of activity or expression of the histone deacetylase polypeptide described above or the level of the nucleic acid molecule described above in the sample. If the level is increased relative to a control, then the subject has an increased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and if the level is decreased relative to a control, then the subject has a decreased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In one embodiment, the polypeptide level is assayed using immunohistochemistry techniques. In another embodiment, the nucleic acid molecule level is assayed using *in situ* hybridization techniques.

Compounds and/or polypeptides identified in the above-described screening methods are also part of the present invention.

## DESCRIPTION OF THE FIGURES

FIG. 1 is a schematic representation of the order in which FIGS. 1A-1O should be viewed.

FIGS. 1A-1C show the cDNA sequence of *HDAC9* (SEQ ID NO: 1). The arrows and numbers in the *HDAC9* sequence indicate exons. The boxed portion of the sequence indicates the HDAC domain.

FIGS. 1D-1G show the cDNA sequence of *HDAC9a* (SEQ ID NO: 3). The arrows and numbers in the *HDAC9a* sequence indicate exons. The boxed portion of the sequence indicates the HDAC domain.

FIGS. 1H-1I show the cDNA sequence of *HDRP*( $\Delta$ NLS) (SEQ ID NO:9).

5 FIGS. 1J-1L show the cDNA sequence of *HDAC9*( $\Delta$ NLS) (SEQ ID NO:5).

FIGS. 1M-1O show the cDNA sequence of *HDAC9a*( $\Delta$ NLS) (SEQ ID NO:7).

FIG. 2 is a schematic representation of the order in which FIGS. 2A-2E should be viewed.

10 FIG. 2A shows the amino acid sequence of HDAC9 (SEQ ID NO: 2).

FIG. 2B shows the amino acid sequence of HDAC9a (SEQ ID NO: 4).

FIG. 2C shows the amino acid sequence of HDAC9( $\Delta$ NLS) (SEQ ID NO: 6).

FIG. 2D shows the amino acid sequence of HDAC9a( $\Delta$ NLS) (SEQ ID NO: 8).

15 FIG. 2E shows the amino acid sequence of and HDRP( $\Delta$ NLS) (SEQ ID NO: 10).

FIG. 3 is a schematic representation of the order in which FIGS. 3A-3C should be viewed.

20 FIGS. 3A-3C show an amino acid sequence alignment of HDRP (SEQ ID NO: 11), HDAC9 (SEQ ID NO: 2), HDAC9a (SEQ ID NO: 4), and HDAC4 (SEQ ID NO: 12) polypeptides. Amino acid sequences of HDAC9 (GenBank Accession: AY032737; SEQ ID NO: 2) and HDAC9a (GenBank Accession: AY032738; SEQ ID NO: 4) are aligned with HDRP (GenBank Accession: BAA34464; SEQ ID NO: 11) and HDAC4 (GenBank Accession: NP\_006028; SEQ ID NO: 12). The identical  
25 residues in all proteins are boxed with solid lines. The similar residues are boxed with dotted lines.

FIG. 4 shows a schematic representation of the human *HDAC9* gene structure. The striped boxes represent exons present in isoforms HDRP, HDAC9a, and HDAC9. The lines represent introns. Broken lines are used for larger introns  
30 (with size in base pair on top). The 5' untranslated region cDNA and coding region cDNA are represented here. Exons 1-12 encode a non-catalytic domain of the

polypeptides, and exons 14-21 encode the histone deacetylase catalytic domain of the polypeptides, which provide the polypeptides with deacetylase activity.

FIG. 5 is a schematic representation of the order in which FIGS. 5A-5D should be viewed.

5        FIGS. 5A-5D show the nucleic acid sequence of *HDAC9*, containing all exons expressed in the various isoforms of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* of the present invention (SEQ ID NO:13).

FIG. 6A is a scanned image of a multiple human tissue Northern blot that was probed to determine mRNA expression of *HDAC9* using a cDNA probe that  
10        recognizes both *HDAC9* and *HDAC9a*. The tissues examined are lane 1, heart; lane 2, brain; lane 3, placenta; lane 4, lung; lane 5, liver; lane 6, skeletal muscle; lane 7, kidney; and lane 8, pancreas. Positions of the RNA size marker in kilobases (kb) are indicated to the left of the blot.

FIG. 6B is a scanned image of an electrophoretic gel showing the results of  
15        RT-PCR analyses of mRNA from the same tissues as examined in the Northern blot of FIG. 6A to determine the distribution of *HDAC9* and *HDAC9a* mRNA among these tissues. PCR products were resolved by agarose gel electrophoresis and visualized by ethidium bromide under UV light. A 1-kb DNA ladder was run on both sides of the gel with the size (in kb) indicated on the left. On the right side, the  
20        expected products for *HDAC9* and *HDAC9a* are indicated as 9 and 9a, respectively.

FIG. 7 is a graph of HDAC enzymatic activity of HDAC anti-FLAG-immunoprecipitated proteins isolated from vector control, HDAC9-FLAG, and HDAC9a-FLAG transfected 293T cells, as measured in fluorescence units using *FLUOR DE LYST*<sup>™</sup> as a substrate in the presence or absence of 1 μM TSA. Results  
25        are shown as the mean of three independent assays. The inset is a scanned image of an anti-FLAG Western blot showing the amount of proteins used in the assay. V, Vector control; 9, HDAC9-FLAG; and 9a, HDAC9a-FLAG.

FIG. 8 is a graph of HDAC enzymatic activity of HDAC anti-FLAG-immunoprecipitated proteins isolated from vector control, and HDAC9a-FLAG  
30        (treated with 2 μM SAHA or left untreated) transfected 293T cells, as measured by <sup>3</sup>H-acetic acid released from <sup>3</sup>H-histones in the presence or absence of 2 μM SAHA.

Vector control; HDAC9a, HDAC9a-FLAG; and HDAC9a+, HDAC9a-FLAG + SAHA.

FIG. 9A shows a scanned image of a Western blot of 293T whole cell lysate and anti-FLAG immunoprecipitates from 293T cells transfected with vector, HDAC9-FLAG or HDAC9a-FLAG using antibodies against MEF2 and FLAG. Top panel, anti-MEF2 Western; bottom panel, anti-FLAG Western. L, 293T whole cell lysate; V, vector control IP; 9, HDAC9-FLAG IP; 9a, HDAC9a-FLAG IP.

FIG. 9B is a graph showing the transcription level of p3XMEF2-*Luc* in the presence or absence of pcDNA3 empty vector (-), pCMV-MEF2C, and/or a vector encoding pFLAG-HDAC9 or pFLAG-HDAC9a. p3XMEF2-*Luc* (100 ng) and pRL-TK (5 ng) were transfected into 293T cells with pcDNA3 empty vector (-) or with pCMV-MEF2C (100 ng) (+) along with the indicated amount of pFLAG-HDAC9 or pFLAG-HDAC9a. pFLAG empty vector was used to adjust the DNA to an equal amount in each transfection. The firefly luciferase activity was first normalized to the co-transfected Renilla luciferase activity and the value for MEF2C alone was then set as 1. Results are shown as the mean of three independent transfections +/- standard deviation.

FIG. 10 shows a schematic representation of the HDAC domains of human non-Sir2 family HDACs and HDRP. The boxes represent histone deacetylase (HDAC) domains.

FIG. 11 is a schematic representation of the order in which FIGS. 11A-11F should be viewed.

FIGS. 11A-11F show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9 (VR1) (SEQ ID NO: 14). Lowercase letters are vector backbone, uppercase letters are HDAC9 sequence. "Acc" was added at the beginning of the HDAC9 sequence for translation initiation.

FIG. 12 is a schematic representation of the order in which FIGS. 12-1 through 12-66 should be viewed.

FIGS. 12-1 through 12-66 show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9a (VR2), with restriction enzyme sites indicated (SEQ ID NO: 14).

FIG. 13 is a schematic representation of the order in which FIGS. 13A-13E should be viewed.

FIGS. 13A-13E show the nucleotide sequence of the vector pFLAG-CMV-5b-HDAC9a (VR2) (SEQ ID NO: 15). Lowercase letters are vector backbone,  
5 uppercase letters are HDAC9a sequence. "Acc" was added at the beginning of the HDAC9a sequence for translation initiation.

FIG. 14 is a schematic representation of the order in which FIGS. 14-1 through 14-61 should be viewed.

FIGS. 14-1 through 14-61 show the nucleotide sequence of the vector  
10 pFLAG-CMV-5b-HDAC9a (VR2), with restriction enzyme sites indicated (SEQ ID NO: 15).

#### DETAILED DESCRIPTION OF THE INVENTION

A protein designated HDRP (See Zhou *et al.*, Proc. Natl. Acad. Sci. USA,  
15 97:1056-1061 (2000)) (also called MITR (See Sparrow *et al.*, EMBO J. 18:5085-5098(1999); Zhang *et al.*, J. Biol. Chem., 276:35-39 (2001); and Zhang *et al.*, Proc. Natl. Acad. Sci. USA, 98:7354-7359 (2001)) that is 50% identical to the N-terminal domains of histone deacetylase 4 (HDAC4) and histone deacetylase 5 (HDAC5) was recently identified. The cloning and characterization of a novel histone deacetylase,  
20 *HDAC9*, of which HDRP is an alternatively spliced isoform is described herein. The cDNA sequence of *HDAC9* is shown in FIGS. 1A-1C (SEQ ID NO: 1), and the HDAC9 amino acid sequence is shown in FIG. 2A (SEQ ID NO: 2). In addition to cloning *HDAC9*, other alternatively spliced isoforms of HDAC9, designated as HDAC9a (a polypeptide that is 132 amino acids shorter at the C-terminal end than  
25 HDAC9), and isoforms of HDAC9, HDAC9a, and HDRP polypeptides that lack the nuclear localization signal (NLS) in the N-terminal non-catalytic end of HDAC9, termed HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS), respectively were also identified. The cDNA sequence of *HDAC9a* is shown in FIGS. 1D-1G (SEQ ID NO: 3), and the HDAC9a amino acid sequence is shown in FIG. 2B (SEQ ID  
30 NO: 4). The cDNA sequence of *HDAC9* lacking amino acids encoding an NLS (*HDAC9*( $\Delta$ NLS)) is shown in FIGS. 1J-1L (SEQ ID NO: 5), and the HDAC9 lacking an NLS amino acid sequence is shown in FIG. 2C (SEQ ID NO: 6). The cDNA

sequence of *HDAC9a* encoding a polypeptide lacking an NLS (*HDAC9a*( $\Delta$ NLS)) is shown in FIGS. 1M-1O (SEQ ID NO: 7), and the *HDAC9a* lacking an NLS amino acid sequence is shown in FIG. 2D (SEQ ID NO: 8). The cDNA sequence of *HDRP* encoding a polypeptide lacking an NLS (*HDRP*( $\Delta$ NLS)) is shown in FIGS. 1H-1I  
5 (SEQ ID NO: 9), and the *HDRP* lacking an NLS amino acid sequence is shown in FIG. 2E (SEQ ID NO: 10).

#### POLYPEPTIDES OF THE INVENTION

The present invention features isolated or recombinant *HDAC9* polypeptides,  
10 *HDAC9a* polypeptides, *HDAC9*( $\Delta$ NLS) polypeptides, *HDAC9a*( $\Delta$ NLS) polypeptides, and *HDRP*( $\Delta$ NLS) polypeptides, and fragments, derivatives, and variants thereof, as well as polypeptides encoded by nucleotide sequences described herein (*e.g.*, other variants). As used herein, the term "polypeptide" refers to a polymer of amino acids, and not to a specific length; thus, peptides, oligopeptides,  
15 and proteins are included within the definition of a polypeptide.

As used herein, a polypeptide is said to be "isolated," "substantially pure," or "substantially pure and isolated" when it is substantially free of cellular material, when it is isolated from recombinant or non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. Typically, the  
20 *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) polypeptide is isolated, substantially pure, or substantially pure and isolated when it has a relative increased concentration or activity of *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS), in comparison to total *HDAC* concentration or activity. Preferably the increased activity or concentration of the  
25 *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) is at least 2-fold, more preferably, at least 5-fold, and most preferably, at least 10 fold, in comparison to total *HDAC* concentration or activity. In addition, a polypeptide can be joined to another polypeptide with which it is not normally associated in a cell (*e.g.*, in a "fusion protein") and still be "isolated," "substantially pure," or  
30 "substantially pure and isolated." An isolated, substantially pure, or substantially pure and isolated polypeptide may be obtained, for example, using affinity



purification techniques described herein, as well as other techniques described herein and known to those skilled in the art.

By a "histone deacetylase polypeptide" is meant a polypeptide having histone deacetylase activity, transcription repression activity, and/or the ability to deacetylate other substrates, for example, transcription factors, including p53, CoRest, E2F, GATA-1, TFIIE, and TFIIIF that normally have a nuclear or cytoplasmic location in a cell. A histone deacetylase polypeptide is also a polypeptide whose activity can be inhibited by molecules having HDAC inhibitory activity. These molecules fall into four general classes: 1) short-chain fatty acids (e.g., 4-phenylbutyrate and valproic acid); 2) hydroxamic acids (e.g. SAHA, Pyroxamide, trichostatin A (TSA), oxamflatin and CHAPs, such as, CHAP1 and CHAP 31); 3) cyclic tetrapeptides (Trapoxin A, Apicidin and Depsipeptide (FK-228, also known as FR9011228); 4) benzamides (e.g., MS-275); and other compounds such as Scriptaid. Examples of such compounds can be found in U.S. Patent Nos. 5,369,108, issued on November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, as well as, Yoshida *et al.*, Bioessays 17, 423-430 (1995), Saito *et al.*, PNAS USA 96, 4592-4597, (1999), Furamai *et al.*, PNAS USA 98 (1), 87-92 (2001), Komatsu *et al.*, Cancer Res. 61(11), 4459-4466 (2001), Su *et al.*, Cancer Res. 60, 3137-3142 (2000), Lee *et al.*, Cancer Res. 61(3), 931-934 and Suzuki *et al.* J. Med. Chem. 42(15), 3001-3003 (1999) the entire content of all of which are hereby incorporated by reference. Examples of such histone deacetylase polypeptides include HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), HDRP( $\Delta$ NLS); a substantially pure polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and a polypeptide having preferably at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, as determined using the BLAST program and parameters described herein.

In one embodiment, the histone deacetylase polypeptide has histone deacetylase activity, transcription repression activity, the ability to deacetylate substrates, or is inhibited by trichostatin A or a hybrid polar compound such as

SAHA. In another embodiment, the HDAC9( $\Delta$ NLS) polypeptide has any two of the above biological activities. In still another embodiment, the HDAC9( $\Delta$ NLS) polypeptide has any three of the above biological activities. In yet another embodiment, the HDAC9( $\Delta$ NLS) polypeptide has all of the above biological activities.

An HDAC9 polypeptide is a histone deacetylase polypeptide as described above. An HDAC9 polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 2, as determined using the BLAST program and parameters described herein.

10 An HDAC9 polypeptide is also a polypeptide that comprises the amino acids encoded by exons 23, 24, 25 and/or 26, and that does not comprise the amino acids encoded by exon 13 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1A-1C, FIG. 4, and FIGS. 5A-5D. Preferably, an HDAC9 polypeptide comprises the sequence of SEQ ID NO: 2. More preferably, an HDAC9 polypeptide consists of  
15 the sequence of SEQ ID NO: 2. An HDAC polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 1.

An HDAC9a polypeptide is a histone deacetylase polypeptide as described above. An HDAC9a polypeptide preferably has at least 60%, more preferably, 70%,  
20 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 4, as determined using the BLAST program and parameters described herein. An HDAC9a polypeptide is also a polypeptide that comprises the amino acids encoded by exon 22, and that does not comprise the amino acids encoded by exons 13, 23, 24, 25, or 26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1D-  
25 1G, FIG. 4, and FIGS. 5A-5D. Preferably, an HDAC9a polypeptide comprises the sequence of SEQ ID NO: 4. More preferably, an HDAC9a polypeptide consists of the sequence of SEQ ID NO: 4. An HDAC9a polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 3.

30 An HDAC9( $\Delta$ NLS) is a histone deacetylase polypeptide as described above. An HDAC9( $\Delta$ NLS) polypeptide does not comprise a nuclear localization signal (NLS). An HDAC9( $\Delta$ NLS) polypeptide preferably has at least 60%, more

preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 6, as determined using the BLAST program and parameters described herein. An HDAC9( $\Delta$ NLS) polypeptide is also a polypeptide that comprises the amino acids encoded by exons 23, 24, 25, and/or 26, and that does not  
5 comprise the amino acids encoded by exons 7 or 13 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1J-1L, and FIGS. 5A-5D. Preferably, an HDAC9( $\Delta$ NLS) polypeptide comprises the sequence of SEQ ID NO: 6. More preferably, an HDAC9( $\Delta$ NLS) polypeptide consists of the sequence of SEQ ID NO: 6. An HDAC9( $\Delta$ NLS) polypeptide is also a polypeptide comprising the amino acid  
10 sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 5.

An HDAC9a( $\Delta$ NLS) polypeptide is a histone deacetylase polypeptide as described above. An HDAC9a( $\Delta$ NLS) does not comprise a nuclear localization signal (NLS). An HDAC9a( $\Delta$ NLS) polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence  
15 identity to SEQ ID NO: 8, as determined using the BLAST program and parameters described herein. An HDAC9a( $\Delta$ NLS) polypeptide is also a polypeptide that comprises the amino acids encoded by exon 22, and that does not comprise the amino acids encoded by exons 7, 13, 23, 24, 25, or 26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1M-1O, and FIGS. 5A-5D. Preferably, an  
20 HDAC9a( $\Delta$ NLS) polypeptide comprises the sequence of SEQ ID NO: 8. More preferably, an HDAC9a( $\Delta$ NLS) polypeptide consists of the sequence of SEQ ID NO: 8. An HDAC9a( $\Delta$ NLS) polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 7.

An HDRP( $\Delta$ NLS) polypeptide is a histone deacetylase polypeptide as  
25 described above. An HDRP( $\Delta$ NLS) does not comprise a nuclear localization signal (NLS). An HDRP( $\Delta$ NLS) polypeptide preferably has at least 60%, more preferably, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% sequence identity to SEQ ID NO: 10, as determined using the BLAST program and parameters described herein. An HDRP( $\Delta$ NLS) polypeptide is also a polypeptide that does not comprise  
30 the amino acids encoded by exons 7 or 13-26 of the *HDAC9* nucleic acid sequence, as shown in FIGS. 1H-1I and FIGS. 5A-5D. Preferably, an HDRP( $\Delta$ NLS) polypeptide comprises the sequence of SEQ ID NO: 10. More preferably, an

HDRP( $\Delta$ NLS) polypeptide consists of the sequence of SEQ ID NO: 10. An HDRP( $\Delta$ NLS) polypeptide is also a polypeptide comprising the amino acid sequence of the polypeptide encoded by the nucleic acid sequence of SEQ ID NO: 9.

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity. In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When a polypeptide is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the polypeptide preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, a polypeptide of the invention comprises an amino acid sequence encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and complements and portions thereof, (*e.g.*, a complement of any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 or a portion of any one of SEQ ID NO: 1 or SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9).

The polypeptides of the invention also encompass fragments and sequence variants. Variants include a substantially homologous polypeptide encoded by the

same genetic locus in an organism, *i.e.*, an allelic variant, as well as other variants. Variants also encompass polypeptides derived from other genetic loci in an organism, but having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and complements and portions thereof, or having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of nucleotide sequences encoding any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.

10 Variants also include polypeptides substantially homologous or identical to these polypeptides but derived from another organism, *i.e.*, an ortholog. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by chemical synthesis. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that

15 are produced by recombinant methods.

As used herein, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 60-65%, typically at least about 70-75%, more typically at least about 80-85%, and most typically greater than about 90-95% or more homologous or identical. A

20 substantially identical or homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule hybridizing to SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a portion thereof, under stringent conditions as more particularly described herein, or will be encoded by a nucleic acid molecule hybridizing to a nucleic acid sequence encoding SEQ ID

25 NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or portion thereof, under stringent conditions as more particularly described herein.

The percent identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence). The nucleotides or amino

30 acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). In

certain embodiments, the length of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) amino acid or nucleotide sequence aligned for comparison purposes is at least 30%, preferably, at least 40%, more preferably, at least 60%, and even more preferably, at least 70%, 80%, 90%, or 100% of the length of the reference sequence, for example, those sequences provided in FIGS. 1A-1O and 2A-2E. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, Proc. Natl. Acad. Sci. USA, 90:5873-5877 (1993). Such an algorithm is incorporated into the BLASTN and BLASTX programs (version 2.2) as described in Schaffer *et al.*, Nucleic Acids Res., 29:2994-3005 (2001). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (*e.g.*, BLASTN) can be used. See <http://www.ncbi.nlm.nih.gov>, as available on August 10, 2001. In one embodiment, the database searched is a non-redundant (NR) database, and parameters for sequence comparison can be set at: no filters; Expect value of 10; Word Size of 3; the Matrix is BLOSUM62; and Gap Costs have an Existence of 11 and an Extension of 1.

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, CABIOS (1989). Such an algorithm is incorporated into the ALIGN program (version 2.0), which is part of the GCG (Accelrys) sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti, Comput. Appl. Biosci., 10: 3-5 (1994); and FASTA described in Pearson and Lipman, Proc. Natl. Acad. Sci USA, 85: 2444-8 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package (available at <http://www.accelrys.com>, as available on August 31, 2001) using either a Blossom 63 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. In yet another embodiment, the percent

identity between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package (available at <http://www.cgc.com>), using a gap weight of 50 and a length weight of 3.

The invention also encompasses HDAC9, HDAC9a, HDAC9( $\Delta$ NLS),  
5 HDAC9a $\Delta$ NLS, and HDRP( $\Delta$ NLS) polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a $\Delta$ NLS, or HDRP( $\Delta$ NLS) polypeptide encoded by a nucleic acid molecule of the invention. Similarity is determined by conserved amino acid substitution. Such substitutions  
10 are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu;  
15 substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, Science 247: 1306-1310 (1990).

A variant polypeptide can differ in amino acid sequence by one or more  
20 substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Further, variant polypeptides can be fully functional or can lack function in one or more activities, for example, in histone deacetylase activity or transcription repression activity. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in  
25 non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncations or a  
30 substitution, insertion, inversion, or deletion in a critical residue or critical region, such critical regions include the HDAC domains, which provide the polypeptide

with deacetylase activity, as shown in the nucleic acid sequences of FIGS. 1A-1G, as well as in the schematic of FIG. 4.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, Science, 244: 1081-1085 (1989)). The latter procedure introduces a single alanine mutation at each of the residues in the molecule (one mutation per molecule). The resulting mutant molecules are then tested for biological activity *in vitro*. Sites that are critical for polypeptide activity can also be determined by structural analysis, such as crystallization, nuclear magnetic resonance, or photoaffinity labeling (See Smith *et al.*, J. Mol. Biol., 224: 899-904 (1992); and de Vos *et al.* Science, 255: 306-312 (1992)).

The invention also includes HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) polypeptide fragments of the polypeptides of the invention. Fragments can be derived from a polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or from a polypeptide encoded by a nucleic acid molecule comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 or a portion thereof and the complements thereof or other variants. The present invention also encompasses fragments of the variants of the polypeptides described herein. Useful fragments include those that retain one or more of the biological activities of the polypeptide as well as fragments that can be used as an immunogen to generate polypeptide-specific antibodies.

Biologically active fragments (peptides that are, for example, 6, 9, 12, 15, 16, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100, or more amino acids in length) can comprise a domain, segment, or motif, for example, an HDAC domain, that has been identified by analysis of the polypeptide sequence using well-known methods, *e.g.*, signal peptides, extracellular domains, one or more transmembrane segments or loops, ligand binding regions, zinc finger domains, DNA binding domains, acylation sites, glycosylation sites, or phosphorylation sites.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for



expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion polypeptides. These  
5 comprise an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a $\Delta$ NLS, or HDRP( $\Delta$ NLS) polypeptide of the invention operatively linked to a heterologous protein or polypeptide having an amino acid sequence not substantially homologous to the polypeptide. "Operatively linked" indicates that the polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to  
10 the N-terminus or C-terminus of the polypeptide. In one embodiment, the fusion polypeptide does not affect the function of the polypeptide per se. For example, the fusion polypeptide can be a GST-fusion polypeptide in which the polypeptide sequences are fused to the C-terminus of the GST sequences. Other types of fusion polypeptides include, but are not limited to, enzymatic fusion polypeptides, for  
15 example,  $\beta$ -galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, and Ig fusions. Such fusion polypeptides, particularly poly-His fusions, can facilitate the purification of recombinant polypeptide. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a polypeptide can be increased by using a heterologous signal sequence. Therefore, in another  
20 embodiment, the fusion polypeptide contains a heterologous signal sequence at its N-terminus.

EP-A 0464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262).  
25 In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. (See Bennett *et al.*, Journal of Molecular Recognition, 8: 52-58 (1995) and Johanson *et al.*, The Journal of Biological Chemistry, 270,16: 9459-9471 (1995)). Thus, this invention also encompasses soluble fusion polypeptides containing a polypeptide of  
30 the invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE).

- A chimeric or fusion polypeptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional
- 5 techniques including automated DNA synthesizers. Alternatively, PCR amplification of nucleic acid fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive nucleic acid fragments that can subsequently be annealed and re-amplified to generate a chimeric nucleic acid sequence (see Ausubel *et al.*, "Current Protocols in Molecular Biology,"
- 10 John Wiley & Sons, (1998), the entire teachings of which are incorporated by reference herein). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A nucleic acid molecule encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide.
- 15 The substantially pure, isolated, or substantially pure and isolated HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a $\Delta$ NLS, or HDRP( $\Delta$ NLS) polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. In one embodiment, the polypeptide is produced by recombinant DNA techniques.
- 20 For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell, and the polypeptide expressed in the host cell. The polypeptide can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.
- 25 In general, HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a $\Delta$ NLS, and HDRP( $\Delta$ NLS) polypeptides of the present invention can be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using art-recognized methods. The polypeptides of the present invention can be used to raise antibodies or to elicit an immune response. The polypeptides can also
- 30 be used as a reagent, e.g., a labeled reagent, in assays to quantitatively determine levels of the polypeptide or a molecule to which it binds (e.g., a receptor or a ligand) in biological fluids. The polypeptides can also be used as markers for cells or tissues

in which the corresponding polypeptide is preferentially expressed, either constitutively, during tissue differentiation, or in a diseased state. The polypeptides can be used to isolate a corresponding binding agent, and to screen for peptide or small molecule antagonists or agonists of the binding interaction. The polypeptides  
5 of the present invention can also be used as therapeutic agents.

#### NUCLEIC ACID MOLECULES OF THE INVENTION

The present invention also features isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules.

10 By a "histone deacetylase nucleic acid molecule" is meant a nucleic acid molecule that encodes a histone deacetylase polypeptide. Such histone nucleic acids include, for example, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule described in detail herein; an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or  
15 SEQ ID NO: 9; a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2,  
20 SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; a nucleic acid that is hybridizable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; a nucleic acid molecule that is hybridizable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3,  
25 SEQ ID NO: 5, or SEQ ID NO: 7; and an isolated nucleic acid molecule that has at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof.

An *HDAC9* nucleic acid molecule is a nucleic acid molecule that encodes an  
30 *HDAC9* polypeptide. In one embodiment, the *HDAC9* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 1; a complement of an isolated nucleic acid comprising SEQ ID NO: 1;

an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2; a nucleic acid that is hybridizable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 2; a  
5 nucleic acid molecule that is hybridizable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 1, as determined using the BLAST program and parameters described herein. In another  
10 embodiment, the *HDAC9* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 1.

An *HDAC9a* nucleic acid molecule is a nucleic acid molecule that encodes an *HDAC9a* polypeptide. An *HDAC9a* nucleic acid molecule preferably has at least 55% sequence identity to SEQ ID NO: 3. In one embodiment, the *HDAC9a* nucleic  
15 acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 3; a complement of an isolated nucleic acid comprising SEQ ID NO: 3; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 4; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 4; a nucleic acid that is  
20 hybridizable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 4; a nucleic acid molecule that is hybridizable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 3; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity  
25 with SEQ ID NO: 3 or a complement thereof, as determined using the BLAST program and parameters described herein. In another embodiment, the *HDAC9a* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 3.

An *HDAC9(ΔNLS)* nucleic acid molecule is a nucleic acid molecule that encodes an *HDAC9(ΔNLS)* polypeptide. In one embodiment, the *HDAC9(ΔNLS)*  
30 nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 5; a complement of an isolated nucleic acid comprising SEQ ID NO: 5; an isolated nucleic acid encoding a histone deacetylase

polypeptide of SEQ ID NO: 6; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 6; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that encodes SEQ ID NO: 6; a nucleic acid molecule that is hybridizeable under high

5 stringency conditions to a nucleic acid comprising SEQ ID NO: 5; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 5 or a complement thereof, as determined using the BLAST program and parameters described herein. In another embodiment, the

10 *HDAC9(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 5.

An *HDAC9a(ΔNLS)* nucleic acid molecule is a nucleic acid molecule that encodes an *HDAC9a(ΔNLS)* polypeptide. In one embodiment, the *HDAC9a(ΔNLS)* nucleic acid molecule is selected from: a nucleic acid molecule that comprises the

15 nucleic acid sequence of SEQ ID NO: 7; a complement of an isolated nucleic acid comprising SEQ ID NO: 7; an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 8; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 8; a nucleic acid that is hybridizeable under high stringency conditions to a nucleic acid molecule that

20 encodes SEQ ID NO: 8; a nucleic acid molecule that is hybridizeable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 7; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ ID NO: 7 or a complement thereof, as determined using the BLAST

25 program and parameters described herein. In another embodiment, the *HDAC9a(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 7.

An "*HDRP(ΔNLS)* nucleic acid molecule" is a nucleic acid molecule that encodes an *HDRP(ΔNLS)* polypeptide. In one embodiment, the *HDRP(ΔNLS)*

30 nucleic acid molecule is selected from: a nucleic acid molecule that comprises the nucleic acid sequence of SEQ ID NO: 9; a complement of an isolated nucleic acid comprising SEQ ID NO: 9; an isolated nucleic acid encoding a histone deacetylase

polypeptide of SEQ ID NO: 10; a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 10; and an isolated nucleic acid molecule that has preferably, at least 55%, more preferably, 60%, 65%, 70%, 75%, 80%, 85%, or 90%, and most preferably, 95% or 99% sequence identity with SEQ  
5 ID NO: 9 or a complement thereof, as determined using the BLAST program and parameters described herein.. In another embodiment, the *HDRP(ΔNLS)* nucleic acid molecule consists of the nucleic acid sequence of SEQ ID NO: 9.

The isolated nucleic acid molecules of the present invention can be RNA, for example, mRNA, or DNA, such as cDNA and genomic DNA. DNA molecules can  
10 be double-stranded or single-stranded; single stranded RNA or DNA can be either the coding, or sense, strand or the non-coding, or antisense, strand. The nucleic acid molecule can include all or a portion of the coding sequence of the gene and can further comprise additional non-coding sequences such as introns and non-coding 3' and 5' sequences (including regulatory sequences, for example). Additionally, the  
15 nucleic acid molecule can be fused to a marker sequence, for example, a sequence that encodes a polypeptide to assist in isolation or purification of the polypeptide. Such sequences include, but are not limited to, those that encode a glutathione-S-transferase (GST) fusion protein and those that encode a hemagglutinin A (HA) polypeptide marker from influenza.

20 An "isolated," "substantially pure," or "substantially pure and isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids that normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (*e.g.*, as in an RNA or cDNA library). For example, an isolated nucleic acid of the invention may  
25 be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system, or reagent mix. In other  
30 circumstances, the material may be purified to essential homogeneity, for example, as determined by agarose gel electrophoresis or column chromatography such as

HPLC. Preferably, an isolated nucleic acid molecule comprises at least about 50, 80, or 90% (on a molar basis) of all macromolecular species present.

With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules that are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb, or 0.1 kb of nucleotides that flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence that is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector are included in the definition of "isolated" as used herein.

Isolated nucleotide molecules also include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. *In vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention are also encompassed by "isolated" nucleotide sequences. Such isolated nucleotide sequences are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (*e.g.*, from other mammalian species), for gene mapping (*e.g.*, by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (*e.g.*, human tissue), such as by Northern blot analysis.

The present invention also pertains to variant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that are not necessarily found in nature but that encode an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide. Thus, for

example, DNA molecules that comprise a sequence that is different from the naturally-occurring *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleotide sequence but which, due to the degeneracy of the genetic code, encode an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
5 *HDRP(ΔNLS)* polypeptide of the present invention are also the subject of this invention.

The invention also encompasses *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleotide sequences encoding portions (fragments), or encoding variant polypeptides such as analogues or derivatives of an  
10 *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide. Such variants can be naturally-occurring, such as in the case of allelic variation or single nucleotide polymorphisms, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion, and substitution of one or more  
15 nucleotides that can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleotide (and/or resultant amino acid) changes are silent or conserved; that is, they do not alter the characteristics or activity of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*,  
20 *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide. In one preferred embodiment, the nucleotide sequences are fragments that comprise one or more polymorphic microsatellite markers.

Other alterations of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules of the invention can  
25 include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, and carbamates), charged linkages (e.g., phosphorothioates or phosphorodithioates), pendent moieties (e.g., polypeptides), intercalators (e.g., acridine or psoralen), chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids).  
30 Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical



interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also pertains to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that hybridize under  
5 high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (e.g., nucleic acid molecules that specifically hybridize to a nucleotide sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). In one embodiment, the invention includes variants described herein that hybridize under high stringency hybridization  
10 conditions (e.g., for selective hybridization) to a nucleotide sequence comprising a nucleotide sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and the complement of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9. In another embodiment, the invention includes variants described herein that hybridize under high stringency  
15 hybridization conditions (e.g., for selective hybridization) to a nucleotide sequence encoding an amino acid sequence of SEQ ID NO: 2 (*HDAC9*), SEQ ID NO: 4 (*HDAC9a*), SEQ ID NO: 6 (*HDAC9(ΔNLS)*), SEQ ID NO: 8 (*HDAC9a(ΔNLS)*), or SEQ ID NO: 10 (*HDRP(ΔNLS)*). In a preferred embodiment, the variant that hybridizes under high stringency hybridizations encodes a polypeptide that has a  
20 biological activity of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide (e.g., histone deacetylase activity or transcription repression activity).

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (e.g., under high stringency conditions). "Specific hybridization," as  
25 used herein, refers to the ability of a first nucleic acid to hybridize to a second nucleic acid in a manner such that the first nucleic acid does not hybridize to any nucleic acid other than to the second nucleic acid (e.g., when the first nucleic acid has a higher similarity to the second nucleic acid than to any other nucleic acid in a sample wherein the hybridization is to be performed). "Stringency conditions" for  
30 hybridization is a term of art that refers to the incubation and wash conditions, e.g., conditions of temperature and buffer concentration, that permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be

- perfectly (*i.e.*, 100%) complementary to the second, or the first and second may share some degree of complementarity that is less than perfect (*e.g.*, 70%, 75%, 85%, 95%). For example, certain high stringency conditions can be used that distinguish perfectly complementary nucleic acids from those of less
- 5 complementarity. "High stringency conditions," "moderate stringency conditions," and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6.3.6 in *Current Protocols in Molecular Biology* (See Ausubel *et al.*, *supra*, the entire teachings of which are incorporated by reference herein). The exact conditions that determine the stringency of
- 10 hybridization depend not only on ionic strength (*e.g.*, 0.2XSSC or 0.1XSSC), temperature (*e.g.*, room temperature, 42°C or 68°C), and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences, and the frequency of occurrence of
- 15 subsets of that sequence within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules. Typically, conditions are used such that sequences at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95% or
- 20 more identical to each other remain hybridized to one another. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions that will allow a given sequence to hybridize (*e.g.*, selectively) with the most similar sequences in the sample can be determined.
- 25 Exemplary conditions are described in Krause and Aaronson, *Methods in Enzymology*, 200:546-556 (1991). Also, in, Ausubel, *et al.*, *supra*, which describes the determination of washing conditions for moderate or low stringency conditions. Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the
- 30 lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the

sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in  $T_m$  of 17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate, or low stringency, depending on the level of mismatch sought.

5 For example, a low stringency wash can comprise washing in a solution containing 0.2XSSC/0.1% SDS for 10 minutes at room temperature; a moderate stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2XSSC/0.1% SDS for 15 minutes at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1XSSC/0.1%SDS  
10 for 15 minutes at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

15 To determine the percent homology or identity of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of one polypeptide or nucleic acid molecule for optimal alignment with the other polypeptide or nucleic acid molecule). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide  
20 positions are then compared, as described above.

The present invention also provides isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence comprising a nucleotide sequence selected from SEQ ID NO: 1,  
25 SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 and also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence encoding an amino acid sequence selected from SEQ ID NO: 2, SEQ ID NO: 4, SEQ  
30 ID NO: 6, SEQ ID NO: 8, and SEQ ID NO: 10. The nucleic acid fragments of the invention are at least about 15, preferably, at least about 18, 20, 23, or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer

fragments, for example, 30 or more nucleotides in length, that encode antigenic polypeptides described herein are particularly useful, such as for the generation of antibodies as described above.

In a related aspect, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*,  
5 and *HDRP(ΔNLS)* nucleic acid fragments of the invention are used as probes or  
primers in assays such as those described herein. "Probes" or "primers" are  
oligonucleotides that hybridize in a base-specific manner to a complementary strand  
of nucleic acid molecules. Such probes and primers include polypeptide nucleic  
acids, as described in Nielsen *et al.*, Science, 254, 1497-1500 (1991). As also used  
10 herein, the term "primer" in particular refers to a single-stranded oligonucleotide that  
acts as a point of initiation of template-directed DNA synthesis using well-known  
methods (*e.g.*, PCR, LCR) including, but not limited to those described herein.

Typically, a probe or primer comprises a region of nucleotide sequence that  
hybridizes to at least about 15, typically about 20-25, and more typically about 40,  
15 50 or 75, consecutive nucleotides of a nucleic acid molecule comprising a  
contiguous nucleotide sequence selected from: SEQ ID NO: 1, SEQ ID NO: 3, SEQ  
ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, the complement of any of SEQ ID NO: 1,  
SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and a sequence  
encoding an amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6,  
20 SEQ ID NO: 8, or SEQ ID NO: 10.

In preferred embodiments, a probe or primer comprises 100 or fewer  
nucleotides, preferably, from 6 to 50 nucleotides, and more preferably, from 12 to 30  
nucleotides. In other embodiments, the probe or primer is at least 70% identical to  
the contiguous nucleotide sequence or to the complement of the contiguous  
25 nucleotide sequence, preferably, at least 80% identical, more preferably, at least 90%  
identical, even more preferably, at least 95% identical, or even capable of selectively  
hybridizing to the contiguous nucleotide sequence or to the complement of the  
contiguous nucleotide sequence. Often, the probe or primer further comprises a  
label, *e.g.*, radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

30 The nucleic acid molecules of the invention such as those described above  
can be identified and isolated using standard molecular biology techniques and the  
sequence information provided in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5,

SEQ ID NO: 7, SEQ ID NO: 9, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, and /or SEQ ID NO: 10. For example, nucleic acid molecules can be amplified and isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based on one or more of the nucleic acid sequences provided above and/or the complement of those sequences. Or such nucleic acid molecules may be designed based on nucleotide sequences encoding one or more of the amino acid sequences provided in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. See generally PCR Technology: Principles and Applications for DNA Amplification (ed. H.A. Erlich, Freeman Press, NY, NY, (1992); PCR Protocols: A Guide to Methods and Applications (Eds. Innis *et al.*, Academic Press, San Diego, CA, (1990); Mattila *et al.*, Nucleic Acids Res., 19: 4967 (1991); Eckert *et al.*, PCR Methods and Applications, 1: 17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford)); and U.S. Patent No. 4,683,202. The nucleic acid molecules can be amplified using cDNA, mRNA, or genomic DNA as a template, cloned into an appropriate vector and characterized by DNA sequence analysis.

Other suitable amplification methods include the ligase chain reaction (LCR) (See Wu and Wallace, Genomics, 4:560 (1989), Landegren *et al.*, Science, 241:1077 (1988)), transcription amplification (Kwoh *et al.*, Proc. Natl. Acad. Sci. USA, 86:1173 (1989)), and self-sustained sequence replication (See Guatelli *et al.*, Proc. Nat. Acad. Sci. USA, 87:1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, that produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be radiolabeled and used as a probe for screening a cDNA library derived from human cells, mRNA in zap express, ZIPLOX, or other suitable vector. Corresponding clones can be isolated, DNA can be obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art-recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of nucleic acid molecules of the present

invention can be accomplished using well-known methods that are commercially available. See, for example, Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York (1989)); Zyskind *et al.*, Recombinant DNA Laboratory Manual, (Acad. Press, (1988)). Using these or similar methods, the

5 polypeptide and the DNA encoding the polypeptide can be isolated, sequenced, and further characterized.

Antisense nucleic acid molecules of the invention can be designed using the nucleotide sequences of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and/or the complement of any of SEQ ID NO: 1, SEQ ID NO:

10 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 and/or a portion of those sequences, and/or the complement of those portion or sequences, and/or a sequence encoding the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or encoding a portion of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Such antisense nucleic

15 acid molecules can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability

20 of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid molecule can be produced biologically using an expression vector into which a nucleic acid molecule has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid molecule

25 will be of an antisense orientation to a target nucleic acid of interest).

In general, the isolated *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid sequences of the invention can be used as molecular weight markers on Southern blots, and as chromosome markers that are labeled to map related gene positions. The nucleic acid sequences can also be used to compare

30 with endogenous DNA sequences in patients to identify genetic disorders (*e.g.*, a predisposition for or susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease), and as probes, such as to hybridize and

discover related DNA sequences or to subtract out known sequences from a sample. The nucleic acid molecules of the present invention can also be used as therapeutic agents.

By a "cell proliferation disease" is meant a disease that is caused by or results in undesirably high levels of cell division, undesirably low levels of apoptosis, or both. For example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer are all examples of cell proliferation diseases. Myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenous leukemia are also cell proliferation diseases.

By a "cell differentiation disease" is meant a disease that is caused by or results in undesirably low levels of cell differentiation, or by undesirably high levels of cell differentiation. For example, cancers such as lymphoma, leukemia, melanoma, ovarian cancer, breast cancer, pancreatic cancer, prostate cancer, colon cancer, and lung cancer are all examples of cell differentiation diseases. Myeloproliferative disorders, including polycythemia vera, essential thrombocythemia, agnogenic myeloid metaplasia, and chronic myelogenous leukemia are also cell differentiation diseases.

By an "apoptotic disease" is meant a condition in which the apoptotic response is abnormal. This may pertain to a cell or a population of cells that does not undergo cell death under appropriate conditions. For example, normally a cell will die upon exposure to apoptotic-triggering agents, such as chemotherapeutic agents, or ionizing radiation. When, however, a subject has an apoptotic disease, for example, cancer, the cell or a population of cells may not undergo cell death in response to contact with apoptotic-triggering agents. In addition, a subject may have an apoptotic disease when the occurrence of cell death is too low, for example, when the number of proliferating cells exceeds the number of cells undergoing cell death, as occurs in cancer when such cells do not properly differentiate.

An apoptotic disease may also be a condition characterized by the occurrence of undesirably high levels of apoptosis. For example, certain neurodegenerative diseases, including but not limited to Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, multiple sclerosis, restenosis, stroke, and ischemic

brain injury are apoptotic diseases in which neuronal cells undergo undesired cell death.

Other diseases for which the polypeptides and nucleic acid molecules of the present invention may be useful for diagnosing and/or treating include, but are not  
5 limited to Huntington's disease.

The *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleic acid molecules of the present invention can further be used to derive primers for genetic fingerprinting, to raise anti-polypeptide antibodies using DNA immunization techniques, and as an antigen to raise anti-DNA antibodies or  
10 elicit immune responses. Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute  
15 biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample.

In addition, the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, and *HDRP(ΔNLS)* nucleotide sequences of the invention can be used to identify and express recombinant polypeptides for analysis, characterization, or therapeutic use,  
20 or as markers for tissues in which the corresponding polypeptide is expressed, either constitutively, during tissue differentiation, or in diseased states. The nucleic acid sequences can additionally be used as reagents in the screening and/or diagnostic assays described herein, and can also be included as components of kits (e.g., reagent kits) for use in the screening and/or diagnostic assays described herein.

Standard techniques, such as the polymerase chain reaction (PCR) and DNA  
25 hybridization, may be used to clone *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs in other species, for example, mammalian homologs. *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* homologs may be readily identified using low-stringency DNA  
30 hybridization or low-stringency PCR with human *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* probes or primers. Degenerate primers encoding human *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or



HDRP( $\Delta$ NLS) polypeptides may be used to clone *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* homologs by RT-PCR.

- Alternatively, additional *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* homologs can be identified by utilizing
- 5 consensus sequence information for *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* polypeptides to search for similar polypeptides in other species. For example, polypeptide databases for other species can be searched for proteins with the HDAC domains described herein. Candidate polypeptides containing such a motif can then be tested for their *HDAC9*, *HDAC9a*,
- 10 *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* biological activities, using methods described herein.

#### EXPRESSION OF THE NUCLEIC ACID MOLECULES OF THE INVENTION

- Another aspect of the invention pertains to nucleic acid constructs containing
- 15 an *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* nucleic acid molecule, for example, one selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, and the complement of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9 (or portions thereof). Yet another aspect of the invention
- 20 pertains to *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, and *HDRP( $\Delta$ NLS)* nucleic acid constructs containing a nucleic acid molecule encoding the amino acid sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. The constructs comprise a vector (*e.g.*, an expression vector) into which a sequence of the invention has been inserted in a sense or antisense orientation.

- 25 As used herein, the term "vector" or "construct" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral
- 30 genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal

mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in

5 recombinant DNA techniques are often in the form of plasmids. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic

10 acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to

15 mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals).

20 Such regulatory sequences are described, for example, in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory

25 sequences).

It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed and the level of expression of polypeptide desired. The expression vectors of the invention can be introduced into host cells to thereby produce

30 polypeptides, including fusion polypeptides, encoded by nucleic acid molecules as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells, such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel,  
5 *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example, using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms  
10 "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included  
15 within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid molecule of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast, or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells, human 293T cells, HeLa cells, NIH 3T3 cells, and mouse  
20 erythroleukemia (MEL) cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of  
25 art-recognized techniques for introducing a foreign nucleic acid molecule (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

30 For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select

these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin, or methotrexate. Nucleic acid molecules encoding a selectable  
5 marker can be introduced into a host cell on the same vector as the nucleic acid molecule of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

- 10 A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector  
15 encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

- The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is  
20 a fertilized oocyte or an embryonic stem cell into which an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule of the invention has been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into the genome or homologous recombinant animals in which  
25 endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleotide sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity.

- As used herein, a "transgenic animal" is a non-human animal, preferably, a  
30 mammal, more preferably, a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians. A

transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably, a mammal, more preferably, a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

- 10           Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191, and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., (1986)). Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, *Current Opinion in Bio/Technology*, 2:823-829 (1991) and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169. Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.*, 20 *Nature*, 385:810-813 (1997) and PCT Publication Nos. WO 97/07668 and WO 97/07669.

#### ANTIBODIES OF THE INVENTION

- Polyclonal and/or monoclonal antibodies that selectively bind one form of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide but not another form of the polypeptide are also provided. Antibodies are also provided that bind a portion of either the variant or reference HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide that contains the polymorphic site or sites.
- 30           In another aspect, the invention provides antibodies to each of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) polypeptides and polypeptide fragments of the invention, *e.g.*, having an amino acid sequence encoded

by SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or a portion thereof, or having an amino acid sequence encoded by a nucleic acid molecule comprising all or a portion of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or another variant, or portion thereof).

The term "purified antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that selectively binds an antigen. A molecule that selectively binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample that naturally contains the polypeptide. Preferably the antibody is at least 60%, by weight, free from proteins and naturally occurring organic molecules with which it naturally associated. More preferably, the antibody preparation is at least 75% or 90%, and most preferably, 99%, by weight, antibody. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')<sub>2</sub> fragments that can be generated by treating the antibody with an enzyme such as pepsin.

The invention provides polyclonal and monoclonal antibodies that selectively bind to an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition," as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, an HDAC9, HDAC9a, HDAC9(ΔNLS), HDAC9a(ΔNLS), or HDRP(ΔNLS) polypeptide of the invention or fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the

blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction.

At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to  
5 prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature*, 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al.*, *Immunol. Today*, 4:72 (1983)), the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)) or trioma techniques. The  
10 technology for producing hybridomas is well known (see generally *Current Protocols in Immunology*, Coligan *et al.*, (eds.) John Wiley & Sons, Inc., New York, NY (1994)). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to  
15 identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, *Current Protocols in*  
20 *Immunology, supra*; Galfre *et al.*, (1977) *Nature*, 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.*, 54:387-402 (1981)). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

25 Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin  
30 library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage

Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; 5 PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, Bio/Technology, 9:1370-1372 (1991); Hay *et al.*, Hum. Antibod. Hybridomas, 3:81-85 (1992); Huse *et al.*, Science, 246:1275-1281 (1989); and Griffiths *et al.*, EMBO J., 12:725-734 (1993).

10           Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

15           In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDAC9a( $\Delta$ NLS) polypeptide of the invention by standard techniques, such as affinity chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly 20 produced polypeptide expressed in host cells. Moreover, an antibody specific for an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDAC9a( $\Delta$ NLS) polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide.

25           The antibodies of the present invention can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent 30 materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, and acetylcholinesterase; examples of suitable



prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride and phycoerythrin; an example of a luminescent material includes luminol; examples of  
5 bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$ , and  $^3\text{H}$ .

#### DIAGNOSTIC AND SCREENING ASSAYS OF THE INVENTION

The present invention also pertains to diagnostic assays for assessing *HDAC*  
10 *9 HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene expression, or for assessing activity of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptides of the invention. In one embodiment, the assays are used in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a cell proliferation disease,  
15 an apoptotic disease, or a cell differentiation disease, or is at risk for (has a predisposition for or a susceptibility to) developing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The invention also provides for prognostic (or predictive) assays for determining whether an individual is susceptible to developing a cell proliferation disease, an apoptotic disease, or a cell  
20 differentiation disease. For example, mutations in the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of symptoms associated with a cell proliferation disease, an apoptotic disease, or a cell  
25 differentiation disease.

Another aspect of the invention pertains to assays for monitoring the influence of agents, or candidate compounds (*e.g.*, drugs or other agents) on the nucleic acid molecule expression or biological activity of polypeptides of the invention, as well as to assays for identifying candidate compounds that bind to an  
30 *HDAC9*, *HDAC9a* polypeptide, an *HDAC9(ΔNLS)* polypeptide, an *HDAC9a(ΔNLS)* polypeptide, or an *HDRP(ΔNLS)* polypeptide. These and other assays and agents are described in further detail in the following sections.

## DIAGNOSTIC ASSAYS

*HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecules, probes, primers, polypeptides, and antibodies to an *HDAC9*,  
5 an *HDAC9a* protein, an *HDAC9(ΔNLS)* protein, an *HDAC9a(ΔNLS)* protein, or an *HDRP(ΔNLS)* protein can be used in methods of diagnosis of a susceptibility to, or likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, as well as in kits useful for diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

10 In one embodiment of the invention, diagnosis of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease is made by detecting a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. The polymorphism can be a mutation in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, such as the  
15 insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of  
20 one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene, or a change in the expression pattern of the various *HDAC9* isoforms. More than one such mutation may be present in a single nucleic acid  
25 molecule.

Such sequence changes cause a mutation in the polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a  
30 premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease can be a synonymous

mutation in one or more nucleotides (*i.e.*, a mutation that does not result in a change in the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide). Such a polymorphism may alter sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the nucleic acid molecule. HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) that has any of the mutations described above is referred to herein as a "mutant nucleic acid molecule."

In a first method of diagnosing a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used (see Ausubel, *et al.*, *supra*). For example, a biological sample from a test subject (a "test sample") of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having, being susceptible to or predisposed for, or carrying a defect for, a cell proliferation disease, an apoptotic disease, or a cell differentiation disease (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source that contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract, or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) is present, and/or to determine which variant(s) encoded by HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) is present. The presence of the polymorphism or variant(s) can be indicated by hybridization of the gene in the genomic DNA, RNA, or cDNA to a nucleic acid probe. A "nucleic acid probe," as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe can contain at least one polymorphism in HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or contains a nucleic acid encoding a particular variant of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS). The probe can be any of the nucleic acid

molecules described above (*e.g.*, the entire nucleic acid molecule, a fragment, a vector comprising the gene, a probe, or primer, etc.).

To diagnose a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, a hybridization sample is formed

5 by contacting the test sample containing *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, with at least one nucleic acid probe. A preferred probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA or genomic DNA sequences described herein. The nucleic

10 acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250, or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be all or a portion of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ

15 ID NO: 7, SEQ ID NO: 9, or the complement of SEQ ID NO: 1 or SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9; or can be a nucleic acid molecule encoding all or a portion of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10. Other suitable probes for use in the diagnostic assays of the invention are described above (*see. e.g.*, probes and primers discussed under the

20 heading, "Nucleic Acids of the Invention").

The hybridization sample is maintained under conditions that are sufficient to allow specific hybridization of the nucleic acid probe to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. "Specific hybridization," as used herein, indicates exact hybridization (*e.g.*, with no mismatches). Specific

25 hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If

30 specific hybridization occurs between the nucleic acid probe and *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in the test sample, then *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* has the

polymorphism, or is the variant, that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or of the presence of a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is therefore diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In Northern analysis (see Current Protocols in Molecular Biology, Ausubel, *et al.*, *supra*), the hybridization methods described above are used to identify the presence of a polymorphism or of a particular variant, associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or of the presence of a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is therefore diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

For representative examples of use of nucleic acid probes, see, for example, U.S. Patent Nos. 5,288,611 and 4,851,330.

Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T, or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen *et al.*, *Bioconjugate Chemistry*, 5 (1994), American Chemical Society, p. 1 (1994)). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. Hybridization of the PNA probe to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* is diagnostic for a decreased

susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In another method of the invention, mutation analysis by restriction digestion can be used to detect a mutant nucleic acid molecule, or nucleic acid molecules  
5 containing a polymorphism(s), if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (and, if necessary, the flanking sequences) in the test sample of genomic DNA from  
10 the test individual. RFLP analysis is conducted as described (see Current Protocols in Molecular Biology, *supra*). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation or polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and therefore indicates the presence or absence of this decreased susceptibility to a cell  
15 proliferation disease, an apoptotic disease, or a cell differentiation disease.

Sequence analysis can also be used to detect specific polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. A test sample of DNA or RNA is obtained from the test individual. PCR or other appropriate methods can be used to amplify the nucleic acid molecule, and/or its  
20 flanking sequences, if desired. The sequence of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or a fragment of the any of those nucleic acid molecules, or an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* cDNA, or a fragment of any of those cDNAs, or an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA,  
25 or a fragment of any of those mRNAs, is determined, using standard methods. The sequence of the above gene, gene fragment, cDNA, cDNA fragment, mRNA, or mRNA fragment is compared with the known nucleic acid sequence of the nucleic acid molecule, cDNA (*e.g.*, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a nucleic acid sequence encoding the protein of SEQ ID  
30 NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or a fragment thereof) or mRNA, as appropriate. The presence of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* indicates that the

individual has a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Allele-specific oligonucleotides can also be used to detect the presence of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
5 *HDRP(ΔNLS)*, through the use of dot-blot hybridization of amplified oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki *et al.*, Nature (London) 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably  
10 approximately 15-30 base pairs, that specifically hybridizes to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and that contains a polymorphism associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. An allele-specific oligonucleotide probe that is specific for particular polymorphisms in *HDAC9*,  
15 *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be prepared, using standard methods (see Current Protocols in Molecular Biology, *supra*).

To identify polymorphisms in the gene that are associated with a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease a test sample of DNA is obtained from the individual. PCR  
20 can be used to amplify all or a fragment of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and its flanking sequences. The DNA containing the amplified *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (or a fragment of any of those genes) is dot-blotted, using standard methods (see Current Protocols in Molecular Biology, *supra*), and the blot is  
25 contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a polymorphism in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, and is  
30 therefore indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual, can be used to identify polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. For example, in one embodiment, an

5 oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "GENECHIPS™," have been generally described in the art, for example, U.S. Patent No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092.

10 These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods that incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor *et al.*, Science, 251:767-777 (1991), Pirrung *et al.*, U.S. Patent No. 5,143,854; PCT Publication No. WO 90/15070; Fodor *et al.*, PCT Publication No. WO 92/10092,

15 and U.S. Patent No. 5,424,186, the entire teachings of each of which are incorporated by reference herein. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Patent No. 5,384,261, the entire teachings of which are incorporated by reference herein.

Once an oligonucleotide array is prepared, a nucleic acid of interest is

20 hybridized to the array and scanned for polymorphisms. Hybridization and scanning are generally carried out by methods described herein and also in, *e.g.*, Published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Patent No. 5,424,186, the entire teachings of which are incorporated by reference herein. In brief, a target nucleic acid sequence that includes one or more previously identified

25 polymorphic markers is amplified by well known amplification techniques, *e.g.*, PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate

30 conditions. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence



hybridizes. The hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array.

Although primarily described in terms of a single detection block, *e.g.*, for detection of a single polymorphism, arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. In alternate arrangements, it will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

Additional descriptions of the use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patent Nos. 5,858,659 and 5,837,832, the entire teachings of which are incorporated by reference herein.

Other methods of nucleic acid analysis can be used to detect polymorphisms in *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. Representative methods include direct manual sequencing (Church and Gilbert Proc. Natl. Acad. Sci. USA 81: 1991-1995, (1988); Sanger *et al.*, Proc. Natl. Acad. Sci. 74: 5463-5467 (1977); Beavis *et al.*, U.S. Patent No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield *et al.*, Proc. Natl. Acad. Sci. USA 86: 232-236 (1991)), mobility shift analysis (Orita *et al.*, Proc. Natl. Acad. Sci. USA 86: 2766-2770 (1989)), restriction enzyme analysis (Flavell *et al.*, Cell 15: 25 (1978); Geever, *et al.*, Proc. Natl. Acad. Sci. USA 78: 5081 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton *et al.*, Proc. Natl. Acad. Sci. USA 85: 4397-4401 (1985)); RNase protection assays (Myers *et al.*, Science 230: 1242 (1985)); use of polypeptides that recognize nucleotide mismatches, such as *E. coli* mutS protein; and allele-specific PCR.

In another embodiment of the invention, diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease can also be made by examining the level of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid, for example, using in situ

5 hybridization techniques known to one skilled in the art, or by examining the level of expression, activity, and/or composition of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, immunohistochemistry, and immunofluorescence. A test

10 sample from an individual is assessed for the presence of an alteration in the level of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid or in the expression and/or an alteration in composition of the polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or for the presence of a particular variant encoded by *HDAC9*, *HDAC9a*,

15 *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. An alteration in expression of a polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*,

20 *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or an alteration in the qualitative polypeptide expression (*e.g.*, expression of a mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide or variant thereof). In a preferred embodiment, diagnosis of a susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease is made by detecting a particular variant

25 encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or a particular pattern of variants. Preferably, increased levels of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or increased expression or activity of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, relative to a control sample, for example, a sample

30 known not to be associated with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, indicates an increased susceptibility or likelihood that the individual has a cell proliferation disease, an apoptotic disease, or a cell

differentiation disease. Alternatively, decreased levels of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* or decreased expression or activity of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, relative to a control sample, for example, a sample  
5 known not to be associated with a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, indicates a decreased susceptibility or likelihood that the individual has a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Both quantitative and qualitative alterations can also be present. An  
10 “alteration” or “modulation” in the polypeptide expression, activity, or composition, as used herein, refers to an alteration in expression or composition in a test sample, as compared with the expression or composition of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is  
15 from the same type of cells), and is from an individual who is not affected by a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.  
20 Similarly, the presence of one or more different variants in the test sample, or the presence of significantly different amounts of different variants in the test sample, as compared with the control sample, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

It is understood that alterations or modulations in polypeptide expression or  
25 function can occur in varying degrees. For example, an alteration or modulation in expression can be an increase, for example, by at least 1.5-fold to 2-fold, at least 3-fold, or, at least 5-fold, relative to the control. Alternatively, the alteration or modulation in polypeptide expression can be a decrease, for example, by at least 10%, at least 40%, 50%, or 75%, or by at least 90%, relative to the control.

30 Various means of examining expression or composition of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and

immunoassays (*e.g.*, David *et al.*, U.S. Patent No. 4,376,110) such as immunoblotting (see also Ausubel *et al.*, *supra*; particularly chapter 10). For example, in one embodiment, an antibody capable of binding to the polypeptide (*e.g.*, as described above), preferably an antibody with a detectable label, can be used. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')<sub>2</sub>) can be used. The term "labeled," with regard to the antibody, is intended to encompass direct labeling of the antibody by coupling (*i.e.*, physically linking) a detectable substance to the antibody, as well as indirect labeling of the antibody by reacting it with another reagent that is directly labeled. An example of indirect labeling is detection of a primary antibody using a fluorescently labeled secondary antibody.

Western blotting analysis, using an antibody as described above that specifically binds to a mutant HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, or an antibody that specifically binds to a non-mutant HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, or an antibody that specifically binds to a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)*, can be used to identify the presence in a test sample of a particular variant of a polypeptide encoded by a polymorphic or mutant *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)*, or the absence in a test sample of a particular variant or of a polypeptide encoded by a non-polymorphic or non-mutant gene. The presence of a polypeptide encoded by a polymorphic or mutant gene, or the absence of a polypeptide encoded by a non-polymorphic or non-mutant gene, is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, as is the presence (or absence) of particular variants encoded by the *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* nucleic acid molecule.

In one embodiment of this method, the level or amount of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in a test sample is compared with the level or amount of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in a control sample. A level or amount of the polypeptide in the test sample that is higher or

lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, and is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Alternatively, the composition of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in a test sample is compared with the composition of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in a control sample. A difference in the composition of the polypeptide in the test sample, as compared with the composition of the polypeptide in the control sample (*e.g.*, the presence of different variants), is diagnostic for a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample. A difference in the amount or level of the polypeptide in the test sample, compared to the control sample; a difference in composition in the test sample, compared to the control sample; or both a difference in the amount or level, and a difference in the composition, is indicative of a decreased susceptibility to a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

Kits (*e.g.*, reagent kits) useful in the methods of diagnosis comprise components useful in any of the methods described herein, including, for example, hybridization probes or primers as described herein (*e.g.*, labeled probes or primers), reagents for detection of labeled molecules, restriction enzymes (*e.g.*, for RFLP analysis), allele-specific oligonucleotides, antibodies that bind to a mutant or to non-mutant (native) HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, means for amplification of nucleic acids comprising HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS), or means for analyzing the nucleic acid sequence of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS), or for analyzing the amino acid sequence of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, etc.

## SCREENING ASSAYS AND AGENTS IDENTIFIED THEREBY

The invention provides methods (also referred to herein as “screening assays”) for identifying the presence of a nucleotide that hybridizes to a nucleic acid of the invention, as well as for identifying the presence of a polypeptide encoded by a nucleic acid of the invention. In one embodiment, the presence (or absence) of a nucleic acid molecule of interest (*e.g.*, a nucleic acid that has significant homology with a nucleic acid of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*) in a sample can be assessed by contacting the sample with a nucleic acid comprising a nucleic acid of the invention (*e.g.*, a nucleic acid having the sequence of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or the complement thereof, or a nucleic acid encoding an amino acid having the sequence of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10, or a fragment or variant of such nucleic acids), under stringent conditions as described above, and then assessing the sample for the presence (or absence) of hybridization. In a preferred embodiment, high stringency conditions are conditions appropriate for selective hybridization. In another embodiment, a sample containing the nucleic acid molecule of interest is contacted with a nucleic acid containing a contiguous nucleotide sequence (*e.g.*, a primer or a probe as described above) that is at least partially complementary to a part of the nucleic acid molecule of interest (*e.g.*, an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid), and the contacted sample is assessed for the presence or absence of hybridization. In a preferred embodiment, the nucleic acid containing a contiguous nucleotide sequence is completely complementary to a part of the nucleic acid molecule of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*.

In any of the above embodiments, all or a portion of the nucleic acid of interest can be subjected to amplification prior to performing the hybridization.

In another embodiment, the presence (or absence) of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, such as a polypeptide of the invention or a fragment or variant thereof, in a sample can be assessed by contacting the sample with an antibody that specifically binds to the

polypeptide of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) (*e.g.*, an antibody such as those described above), and then assessing the sample for the presence (or absence) of binding of the antibody to the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide.

- 5 In another embodiment, the invention provides methods for identifying agents or compounds (*e.g.*, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) that alter or modulate (*e.g.*, increase or decrease) the activity of the polypeptides described herein, or that otherwise interact with the polypeptides
- 10 herein. For example, such compounds can be compounds or agents that bind to polypeptides described herein (*e.g.*, HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrates or agents); that have a stimulatory or inhibitory effect on, for example, activity of polypeptides of the invention; or that change (*e.g.*, enhance or inhibit) the ability of the polypeptides of the invention to
- 15 interact with HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) binding agents; or that alter post-translational processing of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide (*e.g.*, agents that alter proteolytic processing to direct the polypeptide from where it is normally synthesized to another location in the cell, such as the cell
- 20 surface; or agents that alter proteolytic processing such that more polypeptide is released from the cell, etc.). In one example, the binding agent is a cell proliferation disease binding agent, an apoptotic disease binding agent, or a cell differentiation disease binding agent. As used herein, by a "cell proliferation disease binding agent," an "apoptotic disease binding agent," or a "cell differentiation disease
- 25 binding agent" is meant an agent as described herein that binds to a polypeptide of the present invention and modulates a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The modulation can be an increase or a decrease in the severity or progression of the disease. In addition, a cell proliferation disease binding agent, an apoptotic disease binding agent, or a cell differentiation disease
- 30 binding agent includes an agent that binds to a polypeptide that is upstream (earlier) or downstream (later) of the cell signaling events mediated by a polypeptide of the

present invention, and thereby modulates the overall activity of the signaling pathway; in turn, the disease state is modulated.

The candidate compound can cause an increase in the activity of the polypeptide. For example, the activity of the polypeptide can be increased by at least

5 1.5-fold to 2-fold, at least 3-fold, or, at least 5-fold, relative to the control.

Alternatively, the polypeptide activity can be a decrease, for example, by at least 10%, at least 20%, 40%, 50%, or 75%, or by at least 90%, relative to the control.

In one embodiment, the invention provides assays for screening candidate compounds or test agents to identify compounds that bind to or modulate the activity  
10 of polypeptides described herein (or biologically active portion(s) thereof), as well as agents identifiable by the assays. As used herein, a "candidate compound" or "test agent" is a chemical molecule, be it naturally-occurring or artificially-derived, and includes, for example, peptides, proteins, synthesized molecules, for example, synthetic organic molecules, naturally-occurring molecule, for example, naturally  
15 occurring organic molecules, nucleic acid molecules, and components thereof.

In general, candidate compounds for uses in the present invention may be identified from large libraries of natural products or synthetic (or semi-synthetic) extracts or chemical libraries according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise  
20 source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Accordingly, virtually any number of chemical extracts or compounds can be screened using the exemplary methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as  
25 modification of existing compounds. Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available, e.g., from Brandon Associates (Merrimack, NH) and  
30 Aldrich Chemical (Milwaukee, WI). Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova



(Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, FL), and PharmaMar, U.S.A. (Cambridge, MA). In addition, natural and synthetically produced libraries are generated, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. For example, candidate  
5 compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological  
10 library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, Anticancer Drug Des., 12: 145 (1997)). Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

15 In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their activities should be employed whenever possible.

20 When a crude extract is found to modulate (i.e., stimulate or inhibit) the expression and/or activity of the nucleic acids and or polypeptides of the present invention, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and purification process is the careful characterization and  
25 identification of a chemical entity within the crude extract having an activity that stimulates or inhibits nucleic acid expression, polypeptide expression, or polypeptide biological activity. The same assays described herein for the detection of activities in mixtures of compounds can be used to purify the active component and to test derivatives thereof. Methods of fractionation and purification of such heterogenous  
30 extracts are known in the art. If desired, compounds shown to be useful agents for treatment are chemically modified according to methods known in the art. Compounds identified as being of therapeutic value may be subsequently analyzed

using animal models for diseases in which it is desirable to alter the activity or expression of the nucleic acids or polypeptides of the present invention.

In one embodiment, to identify candidate compounds that alter the biological activity, for example, the enzymatic activity or transcriptional repression activity of  
5 an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, a cell, tissue, cell lysate, tissue lysate, or solution containing or expressing an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SE  
ID NO: 8, SEQ ID NO: 10, or another variant encoded by *HDAC9*, *HDAC9a*,  
10 *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)*), or a fragment or derivative thereof (as described above), can be contacted with a candidate compound to be tested under conditions suitable for enzymatic reaction or transcriptional repression reaction, as described herein.

Alternatively, the polypeptide can be contacted directly with the candidate  
15 compound to be tested. The level (amount) of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) biological activity is assessed (*e.g.*, the level (amount) of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) biological activity is measured, either directly or indirectly), and is compared with the level of biological activity in a control (*i.e.*, the level of activity  
20 of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or active fragment or derivative thereof in the absence of the candidate compound to be tested, or in the presence of the candidate compound vehicle only). If the level of the biological activity in the presence of the candidate compound differs, by an amount that is statistically significant, from the level of the biological  
25 activity in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a compound that alters the biological activity of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide. For example, an increase in the level of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS)  
30 enzymatic or transcriptional repression activity relative to a control, indicates that the candidate compound is a compound that enhances (is an agonist of) HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) activity. Similarly,

- a decrease in the enzymatic level or transcriptional repression level of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) activity relative to a control, indicates that the candidate compound is a compound that inhibits (is an antagonist of) HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or
- 5 HDRP( $\Delta$ NLS) activity. In another embodiment, the level of biological activity of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or derivative or fragment thereof in the presence of the candidate compound to be tested, is compared with a control level that has previously been established. A level of the biological activity in the presence of the candidate
- 10 compound that differs from the control level by an amount that is statistically significant indicates that the compound alters HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) biological activity.

- The present invention also relates to an assay for identifying compounds that alter the expression of an *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or
- 15 *HDRP( $\Delta$ NLS)* nucleic acid molecule (e.g., antisense nucleic acids, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) that alter (e.g., increase or decrease) expression (e.g., transcription or translation) of the nucleic acid molecule or that otherwise interact with the nucleic acids described herein, as well as compounds
- 20 identifiable by the assays. For example, a solution containing a nucleic acid encoding an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide can be contacted with a candidate compound to be tested. The solution can comprise, for example, cells containing the nucleic acid or cell lysate containing the nucleic acid; alternatively, the solution can be another solution
- 25 that comprises elements necessary for transcription/translation of the nucleic acid. Cells not suspended in solution can also be employed, if desired. The level and/or pattern of *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* expression (e.g., the level and/or pattern of mRNA or of protein expressed, such as the level and/or pattern of different variants) is assessed, and is compared with the
- 30 level and/or pattern of expression in a control (i.e., the level and/or pattern of *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* expression in the absence of the candidate compound, or in the presence of the candidate

compound vehicle only). If the level and/or pattern in the presence of the candidate compound differs, by an amount or in a manner that is statistically significant, from the level and/or pattern in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a

5 compound that alters the expression of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. Enhancement of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression indicates that the candidate compound is an agonist of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* activity. Similarly, inhibition of *HDAC9*,

10 *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression indicates that the candidate compound is an antagonist of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* activity. In another embodiment, the level and/or pattern of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide(s) (e.g., different variants) in the presence of the

15 candidate compound to be tested, is compared with a control level and/or pattern that has previously been established. A level and/or pattern in the presence of the candidate compound that differs from the control level and/or pattern by an amount or in a manner that is statistically significant indicates that the candidate compound alters *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*

20 expression.

In another embodiment of the invention, compounds that alter the expression of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule or that otherwise interact with the nucleic acids described herein, can be identified using a cell, cell lysate, or solution containing a nucleic

25 acid encoding the promoter region of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene operably linked to a reporter gene. After contact with a candidate compound to be tested, the level of expression of the reporter gene (e.g., the level of mRNA or of protein expressed) is assessed, and is compared with the level of expression in a control (i.e., the level of the expression

30 of the reporter gene in the absence of the candidate compound, or in the presence of the candidate compound vehicle only). If the level in the presence of the candidate compound differs, by an amount or in a manner that is statistically significant, from

the level in the absence of the candidate compound, or in the presence of the candidate compound vehicle only, then the candidate compound is a compound that alters the expression of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, as indicated by its ability to alter expression of a gene that is

5 operably linked to the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* gene promoter. Enhancement of the expression of the reporter indicates that the compound is an agonist of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* activity. Similarly, inhibition of the expression of the reporter indicates that the compound is an antagonist of *HDAC9*, *HDAC9a*,

10 *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* activity. In another embodiment, the level of expression of the reporter in the presence of the candidate compound to be tested, is compared with a control level that has previously been established. A level in the presence of the candidate compound that differs from the control level by an amount or in a manner that is statistically significant indicates

15 that the candidate compound alters *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression.

Compounds that alter the amounts of different variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (e.g., a compound that enhances activity of a first variant, and that inhibits activity of a second variant),

20 as well as compounds that are agonists of activity of a first variant and antagonists of activity of a second variant, can easily be identified using these methods described above.

In other embodiments of the invention, assays can be used to assess the impact of a candidate compound on the activity of a polypeptide in relation to an

25 *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* substrate, for example, an inhibitor of histone deacetylase activity. These inhibitors fall into four general classes: 1) short-chain fatty acids (e.g., 4-phenylbutyrate and valproic acid); 2) hydroxamic acids (e.g., SAHA, Pyroxamide, trichostatin A (TSA), oxamflatin and CHAPs, such as, CHAP1 and CHAP 31); 3) cyclic tetrapeptides

30 (Trapoxin A, Apicidin and Depsiptide (FK-228, also known as FR9011228); 4) benzamides (e.g., MS-275); and other compounds such as Scriptaid. Examples of such assays and compounds can be found in U.S. Patent Nos. 5,369,108, issued on

November 29, 1994, 5,700,811, issued on December 23, 1997, and 5,773,474, issued on June 30, 1998 to Breslow *et al.*, U.S. Patent Nos. 5,055,608, issued on October 8, 1991, and 5,175,191, issued on December 29, 1992 to Marks *et al.*, as well as, Yoshida *et al.*, *supra*; Saito *et al.*, *supra*; Furamai *et al.*, *supra*; Komatsu *et al.*, *supra*; Su *et al.*, *supra*; Lee *et al.*, *supra* and Suzuki *et al.* *supra*, the entire content of all of which are hereby incorporated by reference.

In one example, a cell or tissue that expresses or contains a compound that interacts with HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) (herein referred to as an "HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate," which can be a polypeptide or other molecule that interacts with HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS)) is contacted with HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) in the presence of a candidate compound, and the ability of the candidate compound to alter the interaction between HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) and the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP ( $\Delta$ NLS) substrate is determined, for example, by assaying activity of the polypeptide. Alternatively, a cell lysate or a solution containing the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate, can be used. A compound that binds to HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate can alter the interaction by interfering with, or enhancing the ability of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) to bind to, associate with, or otherwise interact with the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate.

Determining the ability of the candidate compound to bind to HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate can be accomplished, for example, by coupling the candidate compound with a radioisotope or enzymatic label such that binding of the candidate compound to the polypeptide can be determined by detecting the labeled with  $^{125}\text{I}$ ,  $^{35}\text{S}$ ,  $^{14}\text{C}$ , or  $^3\text{H}$ , either directly or indirectly, and the radioisotope detected by direct counting of

radioemmission or by scintillation counting. Alternatively, candidate compound can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product.

5           It is also within the scope of this invention to determine the ability of a candidate compound to interact with the polypeptide without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a candidate compound with HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS),  
10 HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate without the labeling of either the candidate compound, HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS), or the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate (McConnell *et al.*, (1992) Science, 257: 1906-1912). As used herein, a "microphysiometer" (*e.g.*, CYTOSENSOR™) is an analytical  
15 instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and polypeptide.

          In another embodiment of the invention, assays can be used to identify polypeptides that interact with one or more HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptides, as described herein. For example,  
20 a yeast two-hybrid system such as that described by Fields and Song (Fields and Song, Nature 340: 245-246 (1989)) can be used to identify polypeptides that interact with one or more HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptides. In such a yeast two-hybrid system, vectors are  
25 constructed based on the flexibility of a transcription factor that has two functional domains (a DNA binding domain and a transcription activation domain). If the two domains are separated but fused to two different proteins that interact with one another, transcriptional activation can be achieved, and transcription of specific markers (*e.g.*, nutritional markers such as His and Ade, or color markers such as  
30 lacZ) can be used to identify the presence of interaction and transcriptional activation. For example, in the methods of the invention, a first vector is used that includes a nucleic acid encoding a DNA binding domain and an HDAC9, HDAC9a,

HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, variant, or fragment or derivative thereof, and a second vector is used that includes a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a polypeptide that potentially may interact with the HDAC9, HDAC9a,

5 HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, variant, or fragment or derivative thereof (*e.g.*, an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide substrate or receptor). Incubation of yeast containing the first vector and the second vector under appropriate conditions (*e.g.*, mating conditions such as used in the MATCHMAKER™ system

10 from Clontech) allows identification of colonies that express the markers of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS). These colonies can be examined to identify the polypeptide(s) that interact with the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or fragment or derivative thereof. Such polypeptides may be useful as

15 compounds that alter the activity or expression of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, as described above.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize an HDAC9, HDAC9a,

20 HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, or an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate, or other components of the assay on a solid support, in order to facilitate separation of complexed from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. Binding of a candidate compound to the

25 polypeptide, or interaction of the polypeptide with a substrate in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein (*e.g.*, a glutathione-S-transferase fusion protein) can be provided that adds a domain that

30 allows HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) substrate to be bound to a matrix or other solid support.



In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell, cell lysate, tissue, tissue lysate, or solution containing a nucleic acid encoding HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) is contacted with a candidate compound and the expression of appropriate mRNA or polypeptide (*e.g.*, variant(s)) in the cell, cell lysate, tissue, or tissue lysate, or solution, is determined. The level of expression of appropriate mRNA or polypeptide(s) in the presence of the candidate compound is compared to the level of expression of mRNA or polypeptide(s) in the absence of the candidate compound, or in the presence of the candidate compound vehicle only. The candidate compound can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or polypeptide is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator or enhancer of the mRNA or polypeptide expression. Alternatively, when expression of the mRNA or polypeptide is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the mRNA or polypeptide expression. The level of mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting mRNA or polypeptide.

This invention further pertains to novel compounds identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use a compound identified as described herein in an appropriate animal model. For example, a compound identified as described herein (*e.g.*, a candidate compound that is a modulating compound such as an antisense nucleic acid molecule, a specific antibody, or a polypeptide substrate) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such a compound. Alternatively, a compound identified as described herein can be used in an animal model to determine the mechanism of action of such a compound. Furthermore, this invention pertains to uses of novel compounds identified by the above-described screening assays for treatments as described herein. In addition, a compound identified as described herein can be used to alter activity of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, or to

alter expression of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, by contacting the polypeptide or the nucleic acid molecule (or contacting a cell comprising the polypeptide or the nucleic acid molecule) with the compound identified as described herein.

5

#### PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleotides encoding the polypeptides described herein; comprising polypeptides described herein (*e.g.*, SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO:10, and/or other variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*); and/or comprising a compound that alters (*e.g.*, increases or decreases) *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression or *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide activity as described herein. For instance, a polypeptide, protein, fragment, fusion protein or prodrug thereof, or a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention, a compound that alters *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide activity, a compound that alters *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid expression, or an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* substrate or binding partner, can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

25        Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (*e.g.*, NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, *e.g.*, lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic

30

pressure, buffers, coloring, flavoring and/or aromatic substances and the like that do not deleteriously react with the active compounds.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid  
5 solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrrolidone, sodium saccharine, cellulose, magnesium carbonate,  
10 etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, topical, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable  
15 devices, particle acceleration devices ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other compounds.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human  
20 beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free  
25 concentrate in a hermetically sealed container such as an ampule or sachette indicating the quantity of active compound. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampule of sterile water for injection or saline can be  
30 provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a

dynamic viscosity preferably greater than water, can be employed. Suitable formulations include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., that are, if desired, sterilized or mixed with auxiliary agents, *e.g.*, preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The compound may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, *e.g.*, pressurized air.

Compounds described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The compounds are administered in a therapeutically effective amount. The amount of compounds that will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the symptoms of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, that notice

reflects approval by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (*e.g.*, separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the compounds can be separated, mixed together in any combination, present in a single vial or tablet. Compounds assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each compound and administered in FDA approved dosages in standard time courses.

#### METHODS OF THERAPY

The present invention also pertains to methods of treatment (prophylactic, diagnostic, and/or therapeutic) for a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, using an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound. An "HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound" is a compound that alters (*e.g.*, enhances or inhibits) HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide activity and/or *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) nucleic acid molecule expression, as described herein (*e.g.*, an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) agonist or antagonist). therapeutic compounds can alter HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide activity or nucleic acid molecule expression by a variety of means, such as, for example, by providing additional HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or by upregulating the transcription or translation of the *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) nucleic acid molecule; by altering post-translational processing of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide; by altering

transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* variants; or by interfering with *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide activity (e.g., by binding to an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide), or by downregulating the transcription or translation of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid molecule. Representative *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* therapeutic compounds include the following: nucleic acids or fragments or derivatives thereof described herein, particularly nucleotides encoding the polypeptides described herein and vectors comprising such nucleic acids (e.g., a nucleic acid molecule, cDNA, and/or RNA, such as a nucleic acid encoding an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide or active fragment or derivative thereof, or an oligonucleotide; for example, SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or a nucleic acid encoding SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10, or fragments or derivatives thereof); polypeptides described herein (e.g., SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, SEQ ID NO: 10 and/or other variants encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, or fragments or derivatives thereof); *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* substrates; peptidomimetics; fusion proteins or prodrugs thereof; antibodies (e.g., an antibody to a mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, or an antibody to a non-mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide, or an antibody to a particular variant encoded by *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*, as described above); ribozymes; other small molecules; and other compounds that alter (e.g., enhance or inhibit) *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid expression or polypeptide activity, for example, those compounds identified in the screening methods described herein, or that regulate transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* variants (e.g.,

compounds that affect which variants are expressed, or that affect the amount of each variant that is expressed. More than one HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound can be used concurrently, if desired.

5           The HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound that is a nucleic acid is used in the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. The term, "treatment" as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease,  
10           and also lessening the severity or frequency of symptoms of the disease. The therapy is designed to alter (*e.g.*, inhibit or enhance), replace or supplement activity of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in an individual. For example, an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound can be administered in  
15           order to upregulate or increase the expression or availability of the *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* nucleic acid molecule or of specific variants of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS), or, conversely, to downregulate or decrease the expression or availability of the *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or  
20           *HDRP( $\Delta$ NLS)* nucleic acid molecule or specific variants of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS). Upregulation or increasing expression or availability of a native *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* nucleic acid molecule or of a particular variant could interfere with or compensate for the expression or activity of a defective gene  
25           or another variant; downregulation or decreasing expression or availability of a native *HDAC9*, *HDAC9a*, *HDAC9( $\Delta$ NLS)*, *HDAC9a( $\Delta$ NLS)*, or *HDRP( $\Delta$ NLS)* nucleic acid molecule or of a particular variant could minimize the expression or activity of a defective gene or the particular variant and thereby minimize the impact of the defective gene or the particular variant.

30           The HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) therapeutic compound(s) are administered in a therapeutically effective amount (*i.e.*, an amount that is sufficient to treat the disease, such as by

ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount that will be therapeutically effective in the treatment of a particular individual's disorder or condition will depend on the symptoms and severity of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

In one embodiment, a nucleic acid of the invention (*e.g.*, a nucleic acid encoding an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, such as SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which may optionally comprise at least one polymorphism, or a nucleic acid that encodes an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or a variant, derivative or fragment thereof, such as a nucleic acid encoding the protein of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10) can be used, either alone or in a pharmaceutical composition as described above. For example, HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) or a cDNA encoding an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide. If desired, cells that have been transformed with the gene or cDNA or a vector comprising the gene or cDNA can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells that, in nature, lack native HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) expression and activity, or have mutant HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) expression and activity, or have expression of a disease-associated HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) variant,



can be engineered to express an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide or an active fragment of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide (or a different variant of an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide). In a preferred embodiment, nucleic acid encoding the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide, or an active fragment or derivative thereof, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells in an animal. Other gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (*e.g.*, microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used to introduce the desired nucleic acid molecule into a cell.

Alternatively, in another embodiment of the invention, a nucleic acid of the invention; a nucleic acid complementary to a nucleic acid of the invention; or a portion of such a nucleic acid (*e.g.*, an oligonucleotide as described below), can be used in "antisense" therapy, in which a nucleic acid (*e.g.*, an oligonucleotide) that specifically hybridizes to the RNA and/or genomic DNA of HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) is administered or generated *in situ*. The antisense nucleic acid that specifically hybridizes to the RNA and/or DNA inhibits expression of the HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) nucleic acid molecule, *e.g.*, by inhibiting translation and/or transcription. Binding of the antisense nucleic acid can be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interaction in the major groove of the double helix.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid as described above. When the plasmid is transcribed in the cell, it produces RNA that is complementary to a portion of the mRNA and/or DNA that encodes an HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide. Alternatively, the antisense construct can be an oligonucleotide probe which is generated *ex vivo* and introduced

into cells; it then inhibits expression by hybridizing with the mRNA and/or genomic DNA of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. In one embodiment, the oligonucleotide probes are modified oligonucleotides that are resistant to endogenous nucleases, *e.g.* exonucleases and/or endonucleases, thereby rendering them stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Patent Nos. 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy are also described, for example, by Van der Krol *et al.*, *Biotechniques* 6: 958-976 (1988); and Stein *et al.*, *Cancer Res* 48: 2659-2668 (1988). With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site, *e.g.* between the -10 and +10 regions of an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid sequence, are preferred.

To perform antisense therapy, oligonucleotides (RNA, cDNA or DNA) are designed that are complementary to mRNA encoding an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* polypeptide. The antisense oligonucleotides bind to *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, indicates that a sequence has sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid, as described in detail above. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures.

The oligonucleotides used in antisense therapy can be DNA, RNA, or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotides can be modified at the base moiety, sugar

moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotides can include other appended groups such as peptides (*e.g.* for targeting host cell receptors *in vivo*), or compounds facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, Proc. Natl. Acad. Sci. USA 86: 6553-6556 (1989); Lemaitre *et al.*, Proc. Natl. Acad. Sci. USA 84: 648-652 (1987); PCT International Publication No. W088/09810)) or the blood-brain barrier (see, *e.g.*, PCT International Publication No. W089/10134), or hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.*, BioTechniques 6: 958-976 (1988)) or intercalating agents. (See, *e.g.*, Zon, Pharm. Res. 5: 539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule (*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent).

The antisense molecules are delivered to cells that express *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* *in vivo*. A number of methods can be used for delivering antisense DNA or RNA to cells; *e.g.*, antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (*e.g.*, antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systematically. Alternatively, in a preferred embodiment, a recombinant DNA construct is utilized in which the antisense oligonucleotide is placed under the control of a strong promoter (*e.g.*, pol III or pol II). The use of such a construct to transfect target cells in the patient results in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* transcripts and thereby prevent translation of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art and described above. For example, a plasmid, cosmid, YAC, or viral vector can be used to prepare the recombinant DNA

construct that can be introduced directly into the tissue site. Alternatively, viral vectors can be used that selectively infect the desired tissue, in which case administration may be accomplished by another route (e.g., systemically).

Endogenous *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
5 *HDRP(ΔNLS)* expression can also be reduced by inactivating or “knocking out”  
*HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* nucleic acid  
sequences or their promoters using targeted homologous recombination (e.g., see  
Smithies *et al.*, Nature 317: 230-234 (1985); Thomas and Capecchi, Cell 51:  
503-512 (1987); Thompson *et al.*, Cell 5: 313-321 (1989)). For example, a mutant,  
10 non-functional *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
*HDRP(ΔNLS)* (or a completely unrelated DNA sequence) flanked by DNA  
homologous to the endogenous *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*,  
*HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (either the coding regions or regulatory regions  
of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*) can be  
15 used, with or without a selectable marker and/or a negative selectable marker, to  
transfect cells that express *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
*HDRP(ΔNLS)* *in vivo*. Insertion of the DNA construct, via targeted homologous  
recombination, results in inactivation of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*,  
*HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*. The recombinant DNA constructs can be  
20 directly administered or targeted to the required site *in vivo* using appropriate  
vectors, as described above. Alternatively, expression of non-mutant *HDAC9*,  
*HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* can be increased  
using a similar method: Targeted homologous recombination can be used to insert a  
DNA construct comprising a non-mutant, functional *HDAC9*, *HDAC9a*,  
25 *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (e.g., a gene having SEQ ID  
NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9, which  
may optionally comprise at least one polymorphism), or a portion thereof, in place  
of a mutant *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)*  
in the cell, as described above. In another embodiment, targeted homologous  
30 recombination can be used to insert a DNA construct comprising a nucleic acid that  
encodes an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or  
*HDRP(ΔNLS)* polypeptide variant that differs from that present in the cell.

Alternatively, endogenous *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* (*i.e.*, the *HDAC9*,  
5 *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* promoter and/or enhancers) to form triple helical structures that prevent transcription of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in target cells in the body. (See generally, Helene Anticancer Drug Des., 6(6): 569-84 (1991); Helene *et al.*, Ann. N.Y. Acad. Sci., 660: 27-36 (1992); and Maher, Bioassays 14(12): 807-15  
10 (1992)). Likewise, the antisense constructs described herein, by antagonizing the normal biological activity of one of the *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* proteins, can be used in the manipulation of tissue, *e.g.*, tissue differentiation, both *in vivo* and for *ex vivo* tissue cultures. Furthermore, the antisense techniques (*e.g.*, microinjection of antisense molecules,  
15 or transfection with plasmids whose transcripts are anti-sense with regard to an *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* mRNA or gene sequence) can be used to investigate role of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in developmental events, as well as the normal cellular function of *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*,  
20 *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

In yet another embodiment of the invention, other *HDAC9*, *HDAC9a*, *HDAC9(ΔNLS)*, *HDAC9a(ΔNLS)*, or *HDRP(ΔNLS)* therapeutic compounds as described herein can also be used in the treatment or prevention of a cell  
25 proliferation disease, an apoptotic disease, or a cell differentiation disease. The therapeutic compounds can be delivered in a composition, as described above, or by themselves. They can be administered systemically, or can be targeted to a particular tissue. The therapeutic compounds can be produced by a variety of means, including chemical synthesis; recombinant production; *in vivo* production  
30 (*e.g.*, a transgenic animal, such as U.S. Patent No. 4,873,316 to Meade *et al.*), for example, and can be isolated using standard means such as those described herein.

A combination of any of the above methods of treatment (e.g., administration of non-mutant HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), or HDRP( $\Delta$ NLS) polypeptide in conjunction with antisense therapy targeting mutant *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or  
5 *HDRP*( $\Delta$ NLS) mRNA; administration of a first variant encoded by *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) in conjunction with antisense therapy targeting a second encoded by *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS), can also be used.

In another embodiment, the invention is directed to *HDAC9*, *HDAC9a*,  
10 *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) nucleic acid molecules and *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) polypeptides for use as a medicament in therapy. For example, the nucleic acid molecules or polypeptides of the present invention can be used in the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease. In  
15 addition, the *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) nucleic acid molecules and *HDAC9*, *HDAC9a*, *HDAC9*( $\Delta$ NLS), *HDAC9a*( $\Delta$ NLS), or *HDRP*( $\Delta$ NLS) polypeptides described herein can be used in the manufacture of a medicament for the treatment of a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.

20 The invention will be further described by the following non-limiting examples. The teachings of all publications cited herein are incorporated herein by reference in their entirety.

#### EXEMPLIFICATION

##### 25 *Cloning of cDNA encodes a novel HDAC, designated HDAC9*

*HDAC9* was cloned by PCR and 3' rapid amplification of cDNA ends using primers designed from the sequence of human chromosome 7 whose translated product exhibited 80% identity to the HDAC domain of HDAC4, described in detail as follows.

30 Database analyses indicate that *HDRP* is located on chromosome 7 (7p15-p21). The human genome database (February 2001 release) of GenBank was searched using the human HDAC4 amino acid sequence. The TBLASTN program

was used to identify open reading frames downstream of *HDRP* on chromosome 7 that exhibit significant homology to the HDAC domain of HDAC4. Several fragments whose translated products exhibit over 58% identity were retrieved. Two sense primers (OL486, 5'-CCATGGAAACGGTACCCAGCAGGC-3' (SEQ ID NO: 16) and OL487, 5'-CACTCCATCGCTATGATGAAGGG-3' (SEQ ID NO: 17)) and antisense primers (OL484, 5'-AGTTCCTTCATCATAGCGATGG-3' (SEQ ID NO: 18) and OL485, 5'-AATGTACAGGATGCTGGGGT-3' (SEQ ID NO: 19)) each were designed based upon one of these fragments whose translated products matched amino acids 842-873 of HDAC4. RT-PCR was performed using each of the antisense primers and a sense primer (5'-CCCTTGTAAGCTGGTGGAGTTCCCTT-3' (SEQ ID NO: 20)) from the coding region of *HDRP* and human brain cDNA as a template. PCR was performed in a Biometra TGRADIENT Thermocycler for 30 cycles at 95°C for 20 seconds, 60°C for 20 seconds, and 72°C for 120 seconds.

3'-rapid amplification of cDNA ends was performed using the sense primer OL486 and adaptor primer 1 (Clontech), and marathon-ready cDNA from human brain (Clontech, Palo Alto, CA) according to the manufacturer's instruction. The products were re-amplified using nested sense primer OL487 and adaptor primer 2 (Clontech, Palo Alto, CA). PCR products were cloned into pGEM-T-easy vector (Promega, Madison, WI) and sequenced using an automated DNA sequencer at the DNA Sequencing Core Facility of the Memorial Sloan-Kettering Cancer Center, using DNA sequencing methods known to one of skill in the art.

Two cDNAs were cloned from the above-described methods. One cDNA (SEQ ID NO:1) encodes an HDAC9 protein that is 1011 amino acids in length. The other cDNA (SEQ ID NO: 3) encodes an HDAC9a protein that is 879 amino acids long. The cDNA sequence and amino sequence of *HDAC9* and *HDAC9a* are shown in FIGS. 1A-1G and FIGS. 2A-2B, respectively. Database analyses of these cDNAs against human genomic DNA sequences indicated that these two cDNAs are generated by alternatively splicing. An alignment of HDAC9, HDAC9a, *HDRP*, and HDAC4 is shown in FIGS. 3A-3C.

Each of the HDAC9 and HDAC9a nucleic acid sequences were cloned into the pFLAG-CMV-5b vector (Sigma) in frame with the C-terminal FLAG tag. Only

the coding regions plus three extra base pairs (ACC) of cDNA of the HDAC9 and HDAC9a nucleic acid sequences were included in the constructs. These constructs are referred to herein as HDAC9-FLAG and HDAC9a-FLAG, respectively. These constructs are contained in *E. coli*, and can readily be expressed. For HDAC9, the  
5 insert is 3033 bp and for HDAC9a, the insert size is 2637 bp. Both HDAC9 and HDAC9a can be released with EcoRV and BamHI (whose sites have been incorporated in the primers to obtain HDAC9 and HDAC9a coding cDNA for cloning purpose) restriction enzyme digestion.

The *HDAC9* cDNA sequences from the known 5'-end of *HDRP* cDNA to the  
10 3'-untranslated region cloned in this study cover over 511 kb of genomic DNA on chromosome 7. As shown in FIG. 4, the coding region cDNA of *HDAC9* resides in 23 exons spanning 458 kb of genomic sequence. Exons 21, 22, and 23 are one single exon in HDAC9a, but the middle exon that is numbered exon 22 in FIG. 4, containing an in-frame stop codon, is spliced out in HDAC9. In addition, exons 12  
15 and 13 are a single exon used by HDRP. Exon 13 is spliced as part of an intron in HDAC9 and HDAC9a.

Further analysis revealed that exon 7, which contains a nuclear localization signal (NLS) is alternatively spliced in an HDRP isoform, creating HDRP( $\Delta$ NLS). RT-PCR analyses using primers based on sequences from exon 6 and exon 14  
20 indicate that this alternative splicing event also occurs in *HDAC9* and/or *HDAC9a*. Thus, it is possible that at least 6 proteins can be generated from a single *HDAC9* gene by alternatively splicing of its RNA. The cDNA sequences and amino acid sequences for HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) are shown in FIGS. 1A-1O and 2A-2E, respectively.

25

*HDAC9 mRNA is differentially expressed among human tissues*

The expression of *HDAC9* mRNA was determined by Northern blot analysis using a human multiple tissue Northern blot (Clontech, Palo Alto, CA). Hybridization was performed according to the manufacturer's instruction using  
30 ExPressHyb solution (Clontech, Palo Alto, CA). The  $^{32}$ P-random priming labeled 3'-untranslated region common to both *HDAC9* and *HDAC9a* that shares no significant sequence homology with *HDRP* was used as a probe. Two transcripts at



9.8 and 4.1 kb were detected in all tissues examined (FIG. 6A). The 4.1 kb transcript is shorter than the 4.4 kb *HDRP* transcript (See Zhou, *et al.*, Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)). A third transcript at 1.2 kb was detected in placenta (FIG. 6A). Similar to *HDRP* (See Zhou, X., *et al.*, Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)), high levels of *HDAC9* transcripts were detected in brain and skeletal muscle (FIG. 6A).

The distribution of alternatively spliced mRNA variants among tissues was examined by RT-PCR using primers (OL516 5'-TGTGTCATCGAGCTGGCTTC-3' (SEQ ID NO: 21) and OL517 5'-ATCTTCTGCAAGTGGCTCCA-3' (SEQ ID NO: 22)) spanning the alternatively spliced exon 22 and cDNA panel from the same tissues as the multiple tissue Northern blot. PCR was performed in a Biometra TGRADIENT Thermocycler for 30 cycles at 95°C for 20 seconds, 60°C for 20 seconds, and 72°C for 60 seconds. The expected sizes of PCR products were 680 base pairs for *HDAC9* and 993 base pairs for *HDAC9a*. The ratio of *HDAC9* and *HDAC9a* transcripts differed among tissues (FIG. 6B). In the placenta and kidney, the levels of the two transcripts were about the same (FIG. 6B). In the brain, heart, and pancreas, there were more transcripts of *HDAC9* than *HDAC9a*. In the other tissues examined, there were more *HDAC9a* transcripts than *HDAC9* transcripts (FIG. 6B). Under the conditions tested, *HDAC9* transcripts were undetectable in liver (FIG. 6B). The lung had an *HDAC9* product that was larger than expected and abundant. The lung also had low levels of *HDAC9* transcripts and *HDAC9a* transcripts (FIG. 6B). An additional PCR product was also amplified from cDNA of the pancreas; this product was than the expected products from *HDAC9* and *HDAC9a* (FIG. 6B). The identity of the different sized transcripts is unknown.

#### *HDAC9 and HDAC9a possess histone deacetylase activity*

*HDAC9* was named based on sequence homology to *HDAC4* (FIGS. 3A-3C). To determine whether *HDAC9* and *HDAC9a* possess *HDAC* activity, an *HDAC* enzymatic assay was performed using anti-FLAG immunoprecipitated *HDAC9*-FLAG and *HDAC9a*-FLAG.

C-terminal FLAG-tagged *HDAC9* (*HDAC9*-FLAG) and *HDAC9a* (*HDAC9a*-FLAG) expression vectors were constructed using the pFLAG-CMV-5b

vector (Sigma) and PCR amplified coding regions of HDAC9 and HDAC9a in frame with the FLAG-tag to form pFLAG-CMV-5b-HDAC9 (plasmid VR1) and pFLAG-CMV-5b-HDAC9a (plasmid VR2). All constructs were confirmed by DNA sequencing.

- 5 Transfection of human kidney 293T cells, immunoprecipitation using anti-FLAG M2 Agarose (Sigma), Western blot analyses and dual luciferase assays were performed essentially as previously described by Zhou *et al.* (Proc. Natl. Acad. Sci. USA, 97:1056-1061 (2000)). Briefly, the cells (American Type Culture Collection) were cultured in DME HG medium (GIBCO/BRL) supplemented with 10%  
10 (vol/vol) FBS at 37 °C in a 5% CO<sub>2</sub> atmosphere. Transient transfection was performed by using Lipofectamine (GIBCO/BRL) or Fugene 6 (Roche Molecular Biochemicals) according to the manufacturers' instructions. Cells were harvested 24 to 48 hours after transfection and lysed in IP lysis buffer (50 mM Tris·HCl, pH 7.5/120 mM NaCl/5 mM EDTA/0.5% NP-40) at 5 x 10<sup>7</sup> cells per ml.  
15 Immunoprecipitation with anti-FLAG M2-agarose (Sigma, St. Louis, MO) was performed according to the manufacturer's instructions. Immunoprecipitated proteins were released from the agarose beads by using FLAG-peptide and either used directly for HDAC enzymatic activity assays or resolved on SDS/PAGE for Western blot analyses. Anti-FLAG antibody was purchased from Sigma (St. Louis,  
20 MO). Western blot analyses were performed using standard methods.

- HDAC9 and HDAC9a enzymatic activity were assessed with the HDAC Fluorescent Activity Assay/Drug Discovery Kit-AK-500 (BIOMOL Research Laboratories) using a FLUOR DE LYS<sup>TM</sup> that contains an acetylated lysine side chain as a substrate and immunoprecipitated HDAC9-FLAG and HDAC9a-FLAG  
25 polypeptides according to the manufacturer's instruction and a SPECTRAMax<sup>®</sup> GEMINI XS microplate spectrofluorometer using the SOFTmax<sup>®</sup> PRO system (Molecular Devices) at excitation 355 nm and emission 460 nm with a cut off filter of 455 nm. Briefly, HDAC9-FLAG and HDAC9a-FLAG were incubated with the substrate overnight at room temperature in a 96-well plate. The reaction was  
30 stopped by addition of Fluor De Lys<sup>TM</sup> Developer and samples were read with the fluorometer.

As shown in FIG. 7, both HDAC9-FLAG and HDAC9a-FLAG deacetylated the acetylated lysine of FLUOR DE LYS<sup>TM</sup> and the activity of HDAC9 and HDAC9a was comparable. To examine the activity of HDAC9 and HDAC9a, inhibition studies using TSA were carried out by preincubating HDAC9-FLAG and HDAC9a-FLAG with TSA for 15 minutes at room temperature. The assay was then carried out as stated above. As shown in FIG. 7, TSA inhibited HDAC9 and HDAC9a deacetylase activity. The inset gel in FIG. 7 shows the amount of protein used in the assay. SAHA, a potent HDAC inhibitor (Richon *et al.*, Proc. Natl. Acad. Sci. USA, 95:3003-3007 (1998)) also completely inhibited the histone deacetylase activity of HDAC9-FLAG and HDAC9a-FLAG. The HDAC activity of HDAC9 and HDAC9a was about ten times lower than the deacetylase activity of HDAC4 when comparable amount of protein was used under conditions tested here.

HDAC9 and HDAC9a enzymatic activity was also determined through HDAC enzymatic assays using <sup>3</sup>H-histones isolated from murine erythroleukemia cells as a substrate. This assay was performed essentially as described by Richon *et al.* (Proc. Natl. Acad. Sci. USA, 95:3003-3007 (1998)). Briefly, HDAC9-FLAG and HDAC9a-FLAG were incubated with <sup>3</sup>H-histones overnight at 37°C. The reaction was stopped by the addition of 1M HCl/0.1 acetic acid. Released <sup>3</sup>H-acetic acid was extracted with ethyl acetate and quantified by scintillation counting. For inhibition studies, the immunoprecipitated complexes were preincubated with the different HDAC inhibitors for 30 minutes at 4°C.

As shown in FIG. 8, HDAC9a-FLAG deacetylated <sup>3</sup>H-acetyl-histones. SAHA, a potent HDAC inhibitor also completely inhibited the histone deacetylase activity of HDAC9a-FLAG. TSA also inhibited HDAC9a deacetylase activity. Similar results were obtained when HDAC9 was used as the enzyme source.

#### *HDAC9 and HDAC9a repress MEF2-mediated transcription*

The *Xenopus* homolog of HDRP, MITR, was identified as a MEF2 interacting transcriptional repressor (Sparrow *et al.*, EMBO J. 18:5085-5098(1999)) and mouse HDRP also interacts with and represses MEF2 mediated transcription (Zhang *et al.*, J. Biol. Chem. 276:35-39 (2001)). We first tested whether HDAC9-FLAG and HDAC9a-FLAG interact with MEF2. 293 cells were transfected with

vector, HDAC9-FLAG, or HDAC9a-FLAG. The cells were subsequently lysed and HDAC9-FLAG and HDAC9a-FLAG proteins were immunoprecipitated with anti-FLAG antibodies. Western blot analysis of the immunoprecipitated proteins was carried out, using anti-MEF-2 antibody to probe the blot. As shown in FIG. 9A,  
5 both HDAC9 and HDAC9a interacted with MEF2 in 293T cells.

It was then determined whether HDAC9 and HDAC9a repress MEF2-mediated transcription. This determination was carried out as follows. The p3XMEF2-luciferase reporter gene (100 ng) and the vector pRL-TK (Promega) (5 ng) were co-transfected into 293T cells in the absence (pcDNA3 empty vector) or  
10 presence of MEF2C (100 ng of pCMV-MEF2C). HDAC9-F (1 ng, 10 ng, or 100 ng of pFLAG-HDAC9; pFLAG-HDAC9 and HDAC9-FLAG are different constructs, with the FLAG sequence located at opposite ends of the HDAC9 nucleotide, but are functionally equivalent) or HDAC9a-F (1 ng, 10 ng, or 100 ng of pFLAG-HDAC9a; pFLAG-HDAC9a and HDAC9a-FLAG are different constructs, with the FLAG  
15 sequence located at opposite ends of the HDAC9a nucleotide, but are functionally equivalent) was included in a subset of experimental groups with the MEF2C vector. pFLAG empty vector was used to adjust the DNA to an equal amount in each transfection. The cells were harvested 24 to 36 hours after transfection and the luciferase activities were measured using the Dual-Luciferase™ Reporter Assay  
20 System from Promega according to the manufacturer's instruction. The firefly luciferase activity was first normalized to the co-transfected Renilla luciferase activity (encoded by the pRL-TK vector), and the luciferase activity value for cells transfected with MEF2C alone was set at 1. MEF2C activated transcription over 30 times the basal level of transcription. As shown in FIG. 9B, HDAC9-FLAG and  
25 HDAC9a-FLAG repressed MEF2C mediated transcriptional activation in a dose-dependent manner and completely abolished the activation at the 100 ng dose for both HDAC9 and HDAC9a. The transcriptional repression effect of HDAC9 and HDAC9a on MEF2C mediated transcription was a specific effect since a co-transfected reporter gene for transfection efficiency containing a TK promoter was  
30 not repressed by HDAC9 or HDAC9a.

Described herein is the identification and characterization of a new class II HDAC, designated HDAC9. HDAC9 has several alternatively spliced isoforms,

one of which is the previously identified HDRP (Zhou *et al.*, Proc. Natl. Acad. Sci. USA 97:1056-1061 (2000)). HDAC9 and HDAC9a possess HDAC activity, which appears to have a lower specific enzymatic activity than HDAC4. While not wishing to be bound by any particular theory, it is possible that an essential co-factor is lost during immunoprecipitation or does not exist in 293T cells (for example, metastasis-associated protein 2 is essential for the assembly of a catalytically active HDAC1 (Zhang *et al.*, Genes Dev. 13:1924-1935 (1999)), the substrates used are not its natural substrate, or the FLAG tag which interferes with the folding of the protein.

10           Searching the human genome with the HDAC domain from either HDAC1 or HDAC9 identified a total of 10 HDACs in the presently completed human genome sequence, a number of which are schematically represented in FIG. 10. HDACs 1, 2, 3, 8, 4, 5, 6, 7, 9, and 9a all have HDAC domains. HDRP, which is also schematically depicted in FIG. 10, does not have a catalytic domain.

15           All references described herein are incorporated by reference in their entirety. While this invention has been particularly shown and described with reference to preferred embodiment thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the spirit and scope of the invention as defined by the appended  
20   claims.

## CLAIMS

What is claimed is:

- 5
1. An isolated or recombinant histone deacetylase polypeptide, said polypeptide selected from:
- 10
- a) an isolated or recombinant polypeptide comprising SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10; and
- b) an isolated or recombinant polypeptide having at least 60% sequence identity with any one of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.
- 15
2. The isolated or recombinant histone deacetylase polypeptide of Claim 1, said polypeptide selected from:
- a) a polypeptide consisting of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10.
- 20
3. The isolated or recombinant histone deacetylase polypeptide of Claim 1, wherein said polypeptide is human.
4. An isolated nucleic acid molecule selected from the group:
- 25
- a) an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9;
- b) a complement of an isolated nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9
- c) an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or
- 30
- SEQ ID NO: 10;

- d) a complement of an isolated nucleic acid encoding a histone deacetylase polypeptide of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, or SEQ ID NO: 10;
- e) a nucleic acid that is hybridizable under high stringency conditions to a nucleic acid molecule that encodes any of SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, or SEQ ID NO: 8, or a complement thereof; or
- f) a nucleic acid molecule that is hybridizable under high stringency conditions to a nucleic acid comprising SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; and
- g) an isolated nucleic acid molecule that has at least 55% sequence identity with any one of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9, or a complement thereof.
5. The isolated nucleic acid molecule of Claim 4, said nucleic acid molecule consisting of the nucleic acid molecule selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, and SEQ ID NO: 9.
6. The isolated nucleic acid molecule of Claim 4, wherein said nucleic acid molecule is human.
7. A vector comprising the isolated nucleic acid molecule of Claim 4.
8. A cell comprising the vector of Claim 7.
9. A cell comprising the isolated nucleic acid molecule of Claim 4.
10. A purified antibody that selectively binds a polypeptide of Claim 1.
11. A method of identifying a compound that modulates expression of a nucleic acid molecule of Claim 4, said method comprising the steps of:

- a) contacting said nucleic acid molecule with a candidate compound under conditions suitable for expression; and
- b) assessing the level of expression of said nucleic acid molecule, wherein a candidate compound that increases or decreases expression of said nucleic acid molecule relative to a control is a compound that modulates expression of said nucleic acid molecule.
- 5
12. The method of Claim 11, wherein said method is carried out in a cell or animal.
- 10
13. The method of Claim 11, wherein said method is carried out in a cell free system.
14. A method of identifying a compound that modulates the enzymatic activity of the polypeptide of Claim 1, said method comprising the steps of:
- 15
- a) contacting said polypeptide with a candidate compound under conditions suitable for enzymatic reaction; and
- b) assessing the enzymatic activity level of said polypeptide, wherein a candidate compound that increases or decreases the enzymatic activity level of said polypeptide relative to a control is a compound that modulates the enzymatic activity of said polypeptide.
- 20
15. The method of Claim 14, wherein said method is carried out in a cell or animal.
- 25
16. The method of Claim 14, wherein said method is carried out in a cell free system.
17. The method of Claim 14, wherein said polypeptide is further contacted with a substrate for the polypeptide, and wherein said substrate is selected from the group consisting of a cell proliferation disease binding agent, an
- 30



apoptotic disease binding agent, and a cell differentiation disease binding agent.

18. The method of Claim 17, wherein said candidate compound is an inhibitor.
- 5 19. The method of Claim 17, wherein said candidate compound is an activator.
20. A method of identifying a compound that modulates the transcriptional repression activity of the polypeptide of Claim 1, said method comprising
- 10 the steps of:
- a) contacting said polypeptide with a candidate compound under conditions suitable for a transcriptional repression reaction; and
  - b) assessing the transcriptional repression activity level of said polypeptide,
- 15 wherein a candidate compound that increases or decreases the transcriptional repression activity level of said polypeptide relative to a control is a compound that modulates the transcriptional repression activity of said polypeptide.
- 20 21. The method of Claim 20, wherein said method is carried out in a cell or animal.
22. The method of Claim 20, wherein said method is carried out in a cell free system.
- 25 23. The method of Claim 20, wherein said polypeptide is further contacted with a substrate for the polypeptide, and wherein said substrate is selected from the group consisting of a cell proliferation disease binding agent, an apoptotic disease binding agent, and a cell differentiation disease binding agent.
- 30 24. The method of Claim 23, wherein said candidate compound is an inhibitor.

25. The method of Claim 23, wherein said candidate compound is an activator.
26. A method of identifying a compound that modulates expression of a nucleic acid molecule of Claim 4, said method comprising the steps of:
- 5 a) providing a nucleic acid molecule comprising a promoter region of said nucleic acid of Claim 4 or part of a promoter region of said nucleic acid of Claim 4 operably linked to a reporter gene;
- b) contacting said nucleic acid molecule or with a candidate compound; and
- 10 c) assessing the level of said reporter gene,
- wherein a candidate compound that increases or decreases expression of said reporter gene relative to a control is a compound that modulates expression of said nucleic acid molecule of Claim 4.
- 15 27. The method of Claim 26, wherein said method is carried out in a cell.
28. A method of identifying a polypeptide that interacts with a polypeptide of Claim 1 in a yeast two-hybrid system, said method comprising the steps of:
- 20 a) providing a first nucleic acid vector comprising a nucleic acid molecule encoding a DNA binding domain and said polypeptide of Claim 1;
- b) providing a second nucleic acid vector comprising a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide;
- 25 c) contacting said first nucleic acid vector with said second nucleic acid vector in a yeast two-hybrid system; and
- d) assessing transcriptional activation in said yeast two-hybrid system, wherein an increase in transcriptional activation relative to a control indicates that the test polypeptide is a polypeptide that interacts with said
- 30 polypeptide of Claim 1.
29. A pharmaceutical composition comprising a polypeptide of Claim 1.

30. A method of diagnosing a cell proliferation disease, an apoptotic disease, or a cell differentiation disease in a subject, said method comprising the steps of:
- 5      a) obtaining a sample from said subject; and
- b) assessing the level of activity or expression of said polypeptide of Claim 1 in said sample, or detecting the level of said nucleic acid molecule of Claim 4,
- 10      wherein if said level is increased relative to a control, then said subject has an increased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, and wherein if said level is decreased relative to a control, then said subject has a decreased likelihood of having a cell proliferation disease, an apoptotic disease, or a cell differentiation disease.
- 15    31. The method of Claim 30, wherein said level of activity or expression of said polypeptide of Claim 1 in said sample is measured using immunohistochemical techniques.
- 20    32. The method of Claim 30, wherein said level of said nucleic acid molecule of Claim 4 in said sample is measured using *in situ* hybridization techniques.
- 25    33. A method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, said method comprising administering a compound identified by the method of Claim 14.
- 30    34. A method of treating a cell proliferation disease, an apoptotic disease, or a cell differentiation disease, said method comprising administering a compound identified by the method of Claim 20.

1/173

FIG. 1A
FIG. 1B
FIG. 1C
FIG. 1D
FIG. 1E
FIG. 1F
FIG. 1G
FIG. 1H
FIG. 1I
FIG. 1J
FIG. 1K
FIG. 1L
FIG. 1M
FIG. 1N
FIG. 1O

FIG. 1

HDAC93186 bp Coding 151-3186Exon 1

1 ggggaagaga ggcacagaca cagataggag aagggcacgg gctggagcca cttgcaggac tgagggtttt tgacaacaaa ccctaggcgc ctgaagaact  
2  
101 ctaagccaga tgggttggtt ggacgagagc agctcttggc tcagcaaga ATGCACAGTA TGATCAGCTC AGTGGATGTG AAGTCAGAAG TTCCTGTGGG  
2  
201 CCTGGAGCCC ATCTCACCTT TAGACCTAAG GACAGACCTC AGGATGATGA TGCCCGTGGT GGACCTGTGT GTCCGTGAGA AGCAATTGCA GCAGGAATTA  
3  
301 CTTCTTATCC AGCAGCAGCA ACAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACACG GCAGCACCCAG GCTCAGCTTC  
4  
401 AGGAGCATAT CAAGGAACTT CTAGCCATAA AACAGCAACA AGAATCTTA GAAAGGAGC AGAACTGGA GCAGCAGAGG CAAGAACAGG AAGTAGAGAG  
5  
501 GCATCGCAGA GAACAGCAGC TTCTCTCTCT CAGAGGCAAA GATAGAGGAC GAGAAAGGC AGTGGCAAGT ACAGAACTAA AGCAGAAGCT TCAAGAGTTC  
5  
601 CTACTGACTA AATCAGCAAC GAAAGCACT CCAACTAATG GAAATAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA CACGGGTGCC CACCACACAT  
6  
701 CATTTGGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCCTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG GATGATTTC CCCCTCGAAA  
7  
801 AACTGCCTCT GAGCCCAACT TGAAGGTGG GTCCAGGTTA AAACAGAAAG TGGCAGAGAG GAGAACGAGC CCCTTACTCA GGCGGAAGGA TGGAAATGTT

FIG. 1A

8  
901 GTCACTTCAT TCAAGAAGCG AAGTTTGAG GTGACAGAAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCAGTTTC ACCAAACAAT GGGCCACTG  
9  
1001 GAAGTGTAC TGAANAAGAG ACTTCGGTTT TGGCCCTTAC CCTTCATGCC GAGCAATGG TTTTCACAGCA AGCATTCTA ATTCAATGAAG ATTCCATGAA  
9  
1101 CCTGCTAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGC TTCCCGCAGT GCCATCCAG CTCATGCTT CGAATTCAT CAAAGAAAG  
10  
1201 CAGAAGTGTG AGACGAGAC GCTTAGGCAA GGTGTTCTC TGCCCTGGCA GTATGGAGC AGCATCCGG CATCTCCAG CCACCTCAT GTTACTTTAG  
10  
1301 AGGGAAGCC ACCCAACAGC AGCCACCAGG CTCCTCTGCA GCATTTATTA TTGAAGAAC AATGCGACA GCAAAGCTT CTTGTAGCTG GTGGAGTTC  
3/173  
1401 CTTACATCCT CAGTCTCCT TGGCAACAA AGAGAGAATT TCACTGGCA TTAGAGGTAC CCACAATTG CCCCCTCACA GACCCCTGAA CCGAACCCAGC  
11  
1501 TCCTCACCTT TGCCTCAGAG CAGTTGGCT CAGCTGGTCA TTCAACAGCA ACACAGCAA TTCTTGGAGA AGCAGAAGCA ATACCAGCAG CAGATCCACA  
1601 TGAACAAACT GCTTTGAAA TCTATTGAAC AACGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGTTTCTG GGGGACCAGG CGATGAGGA  
12  
1701 AGACAGAGCG CCTCTAGTG GCAACAGCAC TAGCAGGAC AGCAGTCTT GTCTGGATGA CACACTGGGA CAAGTTGGG CTGTGAAGT CAAGGAGGA  
1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAATGG AATCTGGGA GCAGGTGCT TTTATGCAAC AGCTTTCCT GGAACCCAG CACACAGTG  
9

FIG. 1B

4/173

1901 CGCTCTCTGT GCGCCAAGCT CCGCTGGCTG CGGTTGGCAT GGATGGATTA GAGAAACACC GTCTGTCTC CAGGACTCAC TCTTCCCCTG CTGCTCTGT  
2001 TTTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCTGGC TCTGCAACTG GAATTGCCTA TGACCCCTTG ATGCTGAAAC ACCAGTGGT TTGTGGCAAT  
2101 TCCACCACCC ACCCTGAGCA TGCTGGAGCA ATACAGAGTA TCTGGTCAGC ACTGCAAGAA ACTGGGCTGC TAAATAAATG TGAGCGAATT CAAGTCGAA  
2201 AAGCCAGCT GGAGGAATA CAGCTTGTTC ATTCTGAACA TCACTCACTG TTGTATGGCA CCAACCCCTT GGACGGACAG AAGCTGGACC CCAGGATACT  
2301 CCTAGGTGAT GACTCTCAA AGTTTTTTTCT CTCAATTACCT TGTGGTGGAC TTGGGGTGGC CAGTGACACC ATTTGGAATG AGCTACACTC GTCCGGTGCT  
2401 GCACGCATGG CTGTTGGCTG TGTATCGAG CTGGCTTCCA AAGTGGCCTC AGGAGAGCTG AAGAATGGGT TTGCTGTGTG GAGGCCCTT GGCCATCACG  
2501 CTGAAGAATC CACAGCATG GGGTCTGCT TTTTAAATC AGTTGCAAT ACCGCCAAT ACTTGAGAGA CCAACTAAT ATAAGCAAGA TATTGATTGT  
2601 AGATCTGGAT GTTCACCATG GAAAGGTAC CCAGCAGGCC TTTTATGCTG ACCCCAGCAT CCTGTACAT TCACTCCATC GCTATGATGA AGGGAACCTT  
2701 TTCCCTGGCA GTGGAGCCCC AATGAGGTT GGAACAGGCC TTGGAGAAGG GTACAATATA AATATTGCCT GGACAGGTGG CCTTGATCTT CCCATGGGAG  
2801 ATGTTGAGTA CCTTGAAGCA TTCAGGACCA TCGTGAAGCC TGTGGCCAAA GAGTTTGATC CAGACATGGT CTTAGTATCT GCTGGATTG ATGCAATTGG  
2901 AGGCCACACC CCTCTCTAG GAGGTACAA ATGACGGCA AATGTTTTG GTCAATTGAC GAAGCAATG ATGACATGG CTGAIGCAGC TGTGGTGTG  
3001 GCTCTAGAAG GAGGACATGA TCTGACAGCC ATCTGTGATG CATCAGAAGC CTGTGTAAAT GCGCTTCTAG GAAATGACT GGAGCCACTT GCAGAAGATA  
3101 TTCTCCACCA AAGCCGAAT ATGAATGCTG TTATTTCTTT ACAGAAGATC ATTGAAATTC AAAGTATGTC TTTAAAGTTC TCTTAA

FIG. 1C

HDAC9a 3499 bp (Coding 151-2790)

Exon 1

1 ggggaagaga ggacagaca cagataggag aagggaccg gctggagcca cttgcaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact

101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaaga ATGCACAGTA TGATCAGCTC AGTGGATGTG AAGTCAGAAG TTCCTGTGGG

201 CCTGGAGCCC ATCTCACCCTT TAGACCTAAG GACAGACCCTC AGGATGATGA TGCCCGTGGT GGACCTGTGT GTCCGTGAGA AGCAATGCA GCAGGAATTA

301 CTTCTTATCC AGCAGCAGCA ACAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACAGG GCAGCACCCAG GCTCAGCTTC

401 AGGAGCATAT CAAGGAACCT CTAGCCATTA AACAGCAACA AGAATCTCTA GAAAGGAGC AGAACTGGA GCAGCAGAGG CAAGAACAGG AAGTAGAGAG

501 GCATCGCAGA GAACAGCAGC TTCCTCTCTT CAGAGGCCAA GATAGAGGAC GAGAAAGGC AGTGGCAAGT ACAGAACTAA AGCAGAAGCT TCAAGAGTTC

5/173

FIG. 1D



5  
601 CTACTGAGTA AATCAGCAAC GAAAGACACT CCAACTAATG GAAAAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA CACGGCTGCC CACCACACAT  
6  
701 CATTGGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCTTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG GATGATTTCC CCCTTCGAAA  
7  
801 AACTGCCCTCT GAGCCCAACT TGAAGGTGCG GTCCAGGTTA AAACAGAAAG TGGCAGAGAG GAGAAGCAGC CCCTTACTCA GGCGGAAGGA TGGAAATGTT  
8  
901 GTCACTTCAT TCAAGAAGCG AATGTTTGAG GTGACAGAAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCCAGTTC ACCAAACAAT GGGCCAACTG  
9  
1001 GAAGTGTAC TGAATAATGAG ACTTCGGTTT TGCCCCCTAC CCTTCATGCC GAGCAAAAGG TTTCACAGCA AGCAATCTA ATTCTGAAG ATTCCATGAA  
10  
1101 CCTGTAAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGC TTCCCGCAGT GCCATCCCAG CTCAAATGCTT CGAATTCCT CAAAGAAAAG  
10  
1201 CAGAACTGTG AGACGCAGAC GCTTAGGCAA GTGTTCCTC TGCTGGGCA GTATGGAGGC AGCATCCCGG CATCTTCCAG CCACCTCAT GTTACTTTAG  
10  
1301 AGGGAAGCC ACCCAACAGC AGCCACCAGG CTCTCCTGCA GCATTATTA TTGAAGAAGC AAATGGACA GCAAAGCTT CTGTAGCTG GTGGAGTTCC  
11  
1401 CTTACATCCT CAGTCTCCCT TGGCAACAAA AGAGAGAAAT TCACCTGGCA TTAGAGGTAC CCACAAATG CCCCCTCACA GACCCCTGAA CCGAACCCAG  
11  
1501 TCTGCACCTT TGCCTCAGAG CACGTTGGCT CAGCTGGTCA TTCAACAGCA ACACCAGCA TTCTTGGAGA AGCAGAAGCA ATACCAGCAG CAGATCCACA

FIG. 1E

1601 TGAACAACT GCTTTCGAAA TCTATTGAAC AACTGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGCTTCAG GGGGACCAGG CGATGCAGGA  
12  
1701 AGACAGAGCG CCTCTTAGTG GCAACAGCAC TAGGAGCGAC AGCAGTGCCT GTGTGGATGA CACACTGGGA CAAGTTGGGG CTGTGAAGGT CAAGGAGGAA  
1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAATGG AATCTGGGA GCAGGCTGCT TTTATGCAAC AGCCTTTCCT GGAACCCAGG CACACACGTG  
14  
1901 CGCTCTCTGT GCGCCAAGCT CCGCTGGGTG CCGTTGGCAT GGATGGATTA GAGAAACACC GTCTCGTCTC CAGGACTCAC TCTTCCCTGT CTGCCCTCTGT  
2001 TTTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCCCTGGC TCTGCACTG GAATTGCCCTA TGACCCCTTG ATGCTGAAAC ACCAGTGCCT TTGTGGCAAT  
7/173  
2101 TCCACCACC ACCCTGAGCA TGCTGGAAGA ATACAGAGTA TCTGTGACG ACTGCAAGAA ACTGGGTGC TAAATAATG TGAGCGAAT CAAGTTCGAA  
15  
2201 AAGCCAGCCT GGAGGAATA CAGCTTGTC ATTCGACA TCACTCACCTG TTGTAAGGA CCAACCCCTT GGACGGACAG AAGCTGGACC CCAGGATACT  
16  
2301 CCTAGGAT GACTCTCAA AGTTTTTTC CTCATTACTT TGTGTGGAC TTGGGGTGA CAGTGACACC ATTTGGAATG AGCTACATC GTCCGGTGT  
17  
2401 GCACGCATGG CTGTTGGCTG TGTATCGAG CTGGCTTCCA AAGTGGCCTC AGGAGAGCTG AAGAATGGT TTGCTGTTGT GAGGCCCTT GGCCATCAG  
18  
2501 CTGAAGAATC CACAGCCATG GGGTCTGTCT TTTTAAATC AGTTGCAAT ACCGCCAAT ACTTGAGAGA CCACTAAT ATAAGCAAGA TATTGATTGT  
19  
20

FIG. 1F

21  
 2601 AGATCTGGAT GTTCACCATG GAAACGGTAC CCAGCAGGCC TTTTATGCTG ACCCCAGCAT CCTGTACATT TCACTCCATC GCTATGATGA AGGGAACTTT  
 2701 TTCCCTGGCA GTGGAGCCCC AATGAGGTT CGGTTTATTT CTTTACAGCC CCACTTTTAT TTGTATCTTT CAGGTAATTG CATTGCATGA ttaccocctaa  
 STOP CODON  
 2801 tttttctgtc ctttgctggt gttttaaatt acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa gtgcataacc cagagcactg  
 2901 tttgtcaggg aaggttggc tgatttgatg tgttgtttga tgtttattc aagagctccc atgtgcttgt ttccctctct tcttgcttcc ttccatttgc  
 3001 tctcttctct gccaccgtg gtgtgtcttt ctcttcccag gttggaacag gccttgaga aggtacaat ataaatattg cctggacagg tggccttgat  
 3101 cctcccatgg gagatgttga gtaccttgaa gcattcagga ccactgtgaa gcctgtggcc aaagagtgtg atccagacat ggtcttagta tctgctggat  
 24  
 3201 ttgatgcatt ggaaggccac acccctctctc taggagggtg caaagtgaag gcaaaatggt ttggtcattt gacgaagcaa ttgatgacat tggctgatgg  
 25  
 3301 acgtgtggtg ttggctctag aaggaggaca tgatctcaca gccatctgtg atgcatcaga agcctgtgta aatgcccttc taggaatga gctggagcca  
 26  
 3401 cttgcagaag atattctcca ccaaagcccg aatatgaatg ctgttatttc ttacagaag atcattgaa ttcaaagtat gtctttaag ttctcttaa

FIG. 1G

9/173

>HDRP (deltaNLS)

1	ggggaagaga	ggcacagaca	cagataggag	aagggcaccg	gctggagcca
51	cttgaggag	tgagggtttt	tgacaacaaa	ccctagcagc	ctgaagaact
101	ctaagccaga	tggggtggct	ggacgagagc	agctcttggc	tcagcaaaaga
151	atgcacagta	tgatcagctc	agtggatgtg	aagtcagaag	ttcctgtggg
201	cctggagccc	atctcacctt	tagacctaa	gacagacctc	aggatgatga
251	tgcccgtggt	ggaccctgtt	gtccgtgaga	agcaattgca	gcagggaatta
301	cttcttatcc	agcagcagca	acaaatccag	aagcagcttc	tgatagcaga
351	gtttcagaaa	cagcatgaga	acttgacacg	gcagcaccag	gctcagcttc
401	aggagcatat	caaggaaactt	ctagccataa	aacagcaaca	agaactccta
451	gaaaaggagc	agaaactgga	gcagcagagg	caagaacagg	aagtagagag
501	gcatcgcaga	gaacagcagc	ttcctcctct	cagaggcaaa	gatagaggac
551	gagaaaagggc	agtggcaagt	acagaagtaa	agcagaagct	tcaagagttc
601	ctactgagta	aatcagcaac	gaaagacact	ccaaactaatg	gaaaaaatca
651	ttccgtgagc	cgccatccca	agctctggta	cacggctgcc	caccacacat
701	cattggatca	aagctctcca	ccccttagtg	gaacatctcc	atcctacaag
751	tacacattac	caggagcaca	agatgcaaa	gatgatttcc	cccttcgaaa
801	aactgaatcc	tcagtcagta	gcagtctctc	aggctctggt	cccagttcac
851	caaaccaatgg	gccaaactgga	agtgttactg	aaatgagac	ttcgggttttg
901	ccccctaccc	ctcatgccga	gcaaatgggtt	tcacagcaac	gcatttcta
951	tcatgaagat	tccatgaacc	tgctaagtct	ttatacctct	ccttctttgc
1001	ccaacattac	cttgggggctt	cccgcagtgc	catcccagct	caatgcttcg

FIG. 1H

10/173

```
1051 aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg
1101 tgttcctctg cctgggcagt atggaggcag catccgggca tcttccagcc
1151 accctcatgt tacttttagag ggaagaccac ccaacagcag ccaccaggct
1201 ctctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct
1251 tgtagctggt ggagttccct tacatcctca gtctcccttg gcaacaaaag
1301 agagaatttc acctggcatt agaggtaccc acaaattgcc ccgtcacaga
1351 cccctgaacc gaaccagtc tgacaccttg cctcagagca cgttgggtca
1401 gctgggtcatt caacagcaac accagcaatt ctggagaag cagaagcaat
1451 accagcagca gatccacatg aacaaactgc ttctgaaatc tattgaacaa
1501 ctgaagcaac caggcagtca ccttgaggaa gcagagggaag agcttcaggg
1551 ggaccaggcg atgcagggaag acagagcgcc ctctagtggc aacagcacta
1601 ggagcgacag cagtgcctgt gtggatgaca cactgggaca agtggggct
1651 gtgaagggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca
1701 ggaaatggaa tctggggagc aggtgctttt tatgcaacag gtaataggca
1751 aagatttagc tccaggattt gtaattaaag tcattatctg a
```

FIG. 11

11/173

>HDAC9 (deltaNLS)  
1 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca  
51 ctgagaggac tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact  
101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaaga  
151 atgcacagta tgatcagctc agtggatgtg aagtcagaag ttctgtgtgg  
201 cctggagccc atctcacctt tagacctaa gacagacctc aggatgatga  
251 tgcccggtgt ggacctgtt gtccgtgaga agcaattgca gcaggaatta  
301 ctcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga  
351 gtctcagaaa cagcatgaga acttgacacg gcagcaccag gctcagcttc  
401 aggagcatat caaggaaactt ctaggcataa aacagcaaca agaactccta  
451 gaaaaggagc agaaactgga gcagcagagg caagaacagg aagtagagag  
501 gcacgcgaga gaacagcagc ttctctctct cagaggcaaa gatagaggac  
551 gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc  
601 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca  
651 ttccgtgagc cgccatccca agctctggta cacggctgcc caccacacat  
701 cattggatca aagctctcca ccccttagtg gaacatctcc atcctacaag  
751 tacacattac caggagcaca agatgcaaa gattgatttc cccttcgaaa  
801 aactgaatcc tcagtcagta gcagttctcc aggctctggt ccagttcac  
851 caaacaatgg gccaaactgga agtggttactg aaaatgagac ttcgggtttg  
901 cccctaccc ctcatgccga gcaaatgggtt tcacagcaac gcatttcta  
951 tcatgaagat tccatgaacc tgctaaagtct ttatacctct ccttctttgc  
1001 ccaacattac cttggggctt cccgcagtgc catccagct caatgcttcg  
1051 aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg  
1101 tgttcctctg cctgggcagt atggaggcag catcccggca tcttcagcc

FIG. 1J

12/173

1151 accctcatgt tactttagag gaaaagccac ccaacagcag ccaccaggct-  
1201 ctctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct  
1251 tgtagctggg ggaattccct tacatcctca gtctcccttg gcaacaaaag  
1301 agagaatttc acctggcatt agaggtaccc acaaattgcc ccgtcacaga  
1351 cccctgaacc gaaccagtc tgacaccttg cctcagagca cgttggtctca  
1401 gctgggtcatt caacagcaac accagcaatt ctgggagaag cagaagcaat  
1451 accagcagca gatccacatg aacaaaactgc ttctgaaatc tattgaacaa  
1501 ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg  
1551 ggaccaggcg atgcaggaag acagagcgcc ctctagtggc aacagcacta  
1601 ggagcgacag cagtgcctgt gtggatgaca cactgggaca agttggggct  
1651 gtgaagggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca  
1701 ggaatggaa tctggggagc aggtgccttt tatgcaacag cctttcctgg  
1751 aaccacgca cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg  
1801 gttggcatgg atggattaga gaaacacagt ctcgctcca ggactcactc  
1851 tccccctgct gcctctgttt tacctcacc agcaatggac cgccccctcc  
1901 agcctggctc tgcaactgga attgcctatg accccttgat gctgaaacac  
1951 cagtgcgttt gtggcaattc caccaccac tgcaagaaac tgggctgcta aataaatgtg  
2001 acagagtatc tggtcacgac gccagcctgg gccaaataca gcttgttcat  
2051 agcgaattca aggtcgaaaa gctatggcacc aacccctgg acggacagaa  
2101 tctgaacatc actcactgtt aggtactcc taggtgatga ctctcaaaag ttttttctct  
2151 gctggacccc aggtactcc tgggtggactt ggggtggaca gtgacacccat ttggaatgag  
2201 cattaccttg cgggtgctgc acgcattggc gagagctgaa gaatgggttt tcatcgagct  
2251 ctacactcgt cgggtgctgc gagagctgaa gaatgggttt gctgtgtgta  
2301 ggcttccaaa gtggcctcag gaagaatcca cagccatggg gttctgcttt  
2351 ggccccctgg ccatacagct tgcaattac gcccaaatc ttgagagacc aactaaatat  
2401 ttttaattcag

FIG. 1K

13/173

```
2451 aagcaagata ttgattgtag atctggatgt tcaccatgga aacggtaccc  
2501 agcaggcctt ttatgctgac ccagcatec tgtacatttc actccatcgc  
2551 tatgatgaag ggaacttttt cctggcagt ggagcccaa atgaggttgg  
2601 aacaggcctt ggagaaaggt acaatataaa tattgcctgg acaggtggcc  
2651 ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc  
2701 gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc  
2751 tggatttgat gcattggaag gccacacccc tcctctagga gggtacaaag  
2801 tgacggcaaa atgttttggg catttgacga agcaattgat gacattggct  
2851 gatggacgtg tgggtgttggc tctagaagga ggacatgac tcacagccat  
2901 ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga aatgagctgg  
2951 agccacttgc agaagatat ctccaccaa gccggaatat gaatgctgtt  
3001 atttctttac agaagatcat tgaaatcaa agtatgtctt taaagttctc  
3051 ttaa
```

FIG. 1L



14/173

>HDAC9a (deltaNLS)

```

1  ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca
51  cttgcaggac tgagggtttt tgacaacaaa ccctagcagc ctgaagaact
101 ctaagccaga tggggtggct ggacgagagc agctcttggc tcagcaaaaga
151 atgcacagta tgatcagctc agtggatgtg aagtcagaag ttctgtgtggg
201 cctggagccc atctcacctt tagacctaa tagacacctc aggatgatga
251 tgcccgtagt ggaccctgtt gtccgtgaga agcaattgca gcaggaaatta
301 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga
351 gtttcagaaa cagcatgaga acttgacacg gcagcaccag gctcagcttc
401 aggagcatat caaggaactt ctagccataa aacagcaaca agaactccta
451 gaaaaggagc agaaactgga gcagcagagg caagaacagg aagtagagag
501 gcatcgcaga gaacagcagc ttctctctct cagaggcaaa gatagaggac
551 gagaaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc
601 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca
651 ttccgtgagc cgccatccca agctctggta cagggtgcc caccacacat
701 cattggatca aagctctcca ccccttagtg gaacatctcc atcctacaag
751 tacacattac caggagcaca agatgcaaaag gatgatttc cccttcgaaa
801 aactgaatcc tcagtcagta gcagttctcc aggtctggt ccagttcac
851 caaacaatgg gccaaactgga agtgttactg aaatgagac ttcggttttg
901 cccctaccc ctcatgccga gcaaatgggt tcacagcaac gcattctaata
951 tcatgaagat tccatgaacc tgctaagtct ttatacctct ccttctttgc
1001 ccaacattac cttggggctt ccgcagtgcc catcccagct caatgcttcg
1051 aattcactca aagaaaagca gaagtgtgag acgcagacgc ttagggcaagg
1101 tgttcctctg cctggggcagt atggaggcag catccgggca tcttccagcc
1151 accctcatgt tactttagag ggaaaagccac ccaacagcag ccaccaggct

```

FIG. 1M

15/173

1201 ctccctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct  
1251 tgtagctggt ggagttccct tacatcctca gtctcccttg gcaacaaaag  
1301 agagaatttc acctggcatt agaggtaacc acaaattgcc ccgtcacaga  
1351 cccctgaacc gaaccagtc tgcacctttg cctcagagca cgttggtctca  
1401 gctgggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat  
1451 accagcagca gatccacatg aacaaactgc ttctgaaatc tattgaacaa  
1501 ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg  
1551 ggaccaggcg atgcaggaag acagagcgcc ctctagtggc aacagcacta  
1601 ggagcgacag cagtgcctgt gtggatgaca cactgggaca agttggggct  
1651 gtgaagggtca aggaggaacc agtgacagt gatgaagatg ctcagatcca  
1701 ggaaatggaa tctggggagc aggtgctttt tatgcaacag cctttcctgg  
1751 aaccacgca cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg  
1801 gttggcatgg atggattaga gaaacacgt ctctctcca ggactcactc  
1851 ttcccctgct gcctctgttt tacctcacc agcaatggac cgccccctcc  
1901 agcctggctc tgcaactgga attgcctatg accccttgat gctgaaacac  
1951 cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat  
2001 acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg  
2051 agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat  
2101 tctgaacatc actcactgtt gtatggcacc aaccccctgg acggacagaa  
2151 gctggacccc aggatactcc taggtgatga ctctcaaaag tttttttcct  
2201 cattaccttg tggtagactt ggggtggaca gtgacacccat ttggaatgag  
2251 ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct  
2301 ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga  
2351 ggccccctgg ccatacagct gaagaatcca cagccatggg gtctctgctt  
2401 tttaattcag ttgcaattac cgccaaatc ttgagagacc aactaaatat

FIG. 1N

16/173

2451 aagcaagata ttgattgtag atctggatgt tcaccatgga aacggtaccc  
2501 agcaggcctt ttatgctgac ccagcatcc tgtacatttc actccatcgc  
2551 tatgatgaag ggaacttttt ccctggcagt ggagcccaa atgaggttcg  
2601 gtttatttct ttagagcccc acttttattt gtatctttca ggtaattgca  
2651 ttgcatgatt acccctaatt ttcttgtcct ttgctggtgt ttaaatatc  
2701 acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt  
2751 gcataaacca gagcactgtt tgtcaggaa ggttgggctg attgatgtg  
2801 ttgttttgatg tttatttcaa gagctcccat gtgcttggtt tcctctcttc  
2851 ttgcttttct ccatttgctc tcttctctgc ccaccgtggt gtgtctttct  
2901 ctcccagggt tggaaacaggc ctggagaag ggtacaatat aaatattgcc  
2951 tggacagggt gccctgatcc tcccatggga gatgttgagt accctgaagc  
3001 attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg  
3051 tcttagtata tgctggattt gatgcattgg aaggccacac ccctcctcta  
3101 ggaggggtaca aagtgcggc aaaatgtttt ggtcatttga cgaagcaatt  
3151 gatgacattg gctgatggac gtgtgggtgt ggctctagaa ggaggacatg  
3201 atctcacagc catctgtgat gcatcagaag cctgtgtaaa tgcccttcta  
3251 ggaatgagc tggagccact tgcagaagat attctccacc aaagcccgaa  
3301 tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt  
3351 ctttaaagtt ctcttaa

FIG. 10

FIG. 2A
FIG. 2B
FIG. 2C
FIG. 2D
FIG. 2E

>HDAC9 (1011 amino acids)  
MHSMISSVDVKSEVPVGLPI SPLDLRTDLRMMMPVDPV VREKQLQ QELL LIQQQQQI  
QKQLLIAEFQKHENL TRQHQALQEH IKELLAIKQQQELLEKEQKLEQQRQEVEVERH  
RREQQLPPLRGKDRGRERAVASTE VKQLQEFLLSKSATKDTPTNGKNHSVSRHPKLWY  
TAAHHTSLDQSSPPLSGTSPSYKYTLP GAQDAKDDFFLRKKTASEPNLKVRSRLKQKVAE  
RRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSSPGSPSPNNGPTGSVTENETSVLP  
PTPHAEQMV SQQRILIHEDSMNLLSLYTSPSLPNI TLGLPAVPSQLNASNSLKEKQKCE  
TQTLRQGVPLPGQYGGSI PASSSHPHVTLEGKPPNSSHQALLQHLLLKEQMRQQKLLVA  
GGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSA PLPQSTLAQLVIQQQHQQFL  
EKQKQYQQQIHMNKL LSKSIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDS  
SACVDDTLGQVGAVKVEEPVDSDEDAQIQEMESGEQA AFMQQPFLPTHTRALSVRQA  
PLAAVGMDGLEKHRLVSRTHSSPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCG  
NSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTNP LD  
GQKLDPRILLGDDSQKFFSSLP CGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVAS  
GELKNGFAVVRPPGHHAEE STAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNG  
TQQAFYADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNINIAWTGGLDPPMGDV  
EYLEAFRTIVKPVAKEFPDMVLVSAGFDAL EGHTPPLGGYKVTAKCFGHLTKQLMTLA  
DGRVVLALGGHDLTAICDASEACVNALLGNELEPLAEDILHQSPNMNAVISLQKII EI  
QMSLSLKFS

FIG. 2A

FIG. 2

18/173

>HDAC9a (879 amino acids)  
MHSMISSVDVKSEVPVGLPISPDLRDLRMMMPVDPVVRKQLQQELLIIQQQQQI  
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH  
RREQQLPPLRGKDRGRERAVASTEVEKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLVY  
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTASEPNIKVRSLIKQKVAE  
RRSSPLLRRKDGNVVTSFKKRMFEVTESSVSSSPGSGPSSPNNGPTGSVTENETSVLP  
PTPHAEQMVVSQQRIILIHEDSMNLLSLYTSPSLPNITLGLPAVPSQNLNASNSLKEKQKCE  
TQTLRQGVPLPGQYGGSI PASSSHPHVTLEGKPPNSSHQALLQHLLLEQMRQQLLVA  
GGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQSTLAQLVIQQQHQQFL  
EKQKYOQQIHMNKLKLSKIEQLKQPGSHLEEAEEELQGDQAMQEDRAPSSGNSTRSDS  
SACVDDTLGQVGAVKVEFPVDSDEDAQIQEMESGEQAQFMQQPFLEPTHTRALSVRQA  
PLAAVGMDGLEKHRLVSRTHSSPPAASVLPHPAMDRPLQPGSATGIAYDPLMLKHQCVCG  
NSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLVHSEHHSLLYGTPILD  
GQKLDPRILLGDDSQKFFSSLLPCGGLGVDSDTIWNELHSSGAARMAVGCVIELASKVAS  
GELKNGFVVRPPGHHAEEESTAMGFCFFNSVAITAKYLRDQLNISKILIVDLDVHHGNG  
TQQAIFYADPSILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFYLYLSGNCIA

FIG. 2B

19/173

>HDAC9 (ANLS) (967 amino acids)  
MHSMISSVDVKSEVPVGLPI SPLDLRTDLRMMPVDPVVRKQLQQELLIIQQQQQI  
KQQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEVEVERH  
RREQQLPPLRGKDRGRERAVASTEVKQKLQEFLLSKSATKDTPTNGKNHSVSRHPKLY  
TAAHHTSLDQSSPPLSGTSPSYKYTTLPGAQDAKDDFPLRKTESSVSSSPGSGPSSPNN  
GPTGSVTENETSVLPPTPHAEQMVSQORILIHEDSMNLLSLYTSPSLPNITLGLPAVPS  
QLNASNSLKEKQKCEQTQLRQGVPLPGQYGGSI PASSSHPHVTLEGKPPNSSHQALLQH  
LLLKEQMRQQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS  
TLAQLVIOQQHQQFLEKQKQYQQQIHMNKL LSKSIEQLKQPGSHLEEAEEELQGDQAMQ  
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKVEEPVDSDEDAQIQEMESGEQA AFMQQP  
FLEPTHTRALSVRQAPLA AVGMDGLEKHLVSRTHSSPAASVLPHPAMDRLQPGSATG  
IAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLV  
HSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLP CGGLGVDSDTIWNELHSSGAAR  
MAVGCVIELASKVASGELKNGFAVVRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNI  
SKILIVDLDVHHNGTQQA FYADPSILYISLHRYDEGNFFPGSGAPNEVGTGLGEGYNI  
NIAWTGGLDPPMGDVEYLEAFRTIVKPVAKFDPDMVLVSAGFDALEGHTPPLGGYKVT  
AKCFGHLTKQLMTLADGRVVLALLEGGHDLT AICDASEACVNALLGNELEPLAEDILHQ  
PNMNAVISLQKIIEIQSMSLKFS

FIG. 2C

20/173

>HDAC9a (ANLS) (835 amino acids)  
MHSMISSVDVKSEVPVGLEPISPLDLRTDLRMMPVVDPVREKQLQQELLIIQQQQI  
QKQLLIAEFQKQHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH  
RREQQLPPLRGKDRGRERAVASTEVEKQLQEFFLLSKSATKDTPTNGKNHSVSRHPKLWY  
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTESSVSSSPGSGPSSPNN  
GPTGSVTENETSVLPPTPHAEQMVSQQRILIHEDSMNLLSLYTSFSLPNIITGLPAVPS  
QLNASNSLKEKQKCEQTQLRQGVPLPGQYGGSI PASSSHPHVTLEGKPPNSSHQALLQH  
LLLKEQMRQKLLVAGGVPLHPQSPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS  
TLAQLVIQQQHQQFLEKQKQYQQQIHMNKLKSKSIEQLKQPGSHLEEAEEELQGDQAMQ  
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKKEEPVDSEDAQIQEMESGEQAQAFMQQP  
FLEPTHTRALSVRQAPLAAVGMGDLGKHLVSRTHSSPAAASVLPHPAMDRPLQPGSATG  
IAYDPLMLKHQCVCGNSTTHPEHAGRIQSIWSRLQETGLLNKCERIQGRKASLEEIQLV  
HSEHHSLLYGTNPLDGQKLDPRILLGDDSQKFFSSLPCGGLGVDSDTIWNELHSSGAAR  
MAVGCVIELASKVASGELKNGFAVVRRPPGHHAEESTAMGFCFFNSVAITAKYLRDQLNI  
SKILIVDLVDVHHGNGTQQAIFYADPSILYISLHRYDEGNFFPGSGAPNEVRFISLEPHFY  
LYLSGNCIA

FIG. 2D

21/173

>HDRPa (HDRP ANLS) (546 amino acids)  
MHSMISSVDVKSEVPVGLPI SPLDLRTDLRMMPVDPVVRKQLQQELLLLIQQQQQI  
QKQLLIAEFQKHENLTRQHQAQLQEHIKELLAIKQQQELLEKEQKLEQQRQEQEVERH  
RREQQLPPLRGKDRGRERAVASTEVKQLQEFLLSKSATKDTPTNGKNHVSVRHPKLWY  
TAAHHTSLDQSSPPLSGTSPSYKYTLPGAQDAKDDFPLRKTESSVSSSPGSGPSSPNN  
GPTGSVTENETSVLPPTPHAEQMV SQQRILIHEDSMNLLSLYTSPSLPNTITLGLPAVPS  
QLNASNSLKEKQKCETQTLRQGVPLPGQYCGSIPASSSHPHVTLEGKPPNSSHQALLQH  
LLLKEQMRQKLLVAGGVPLHPQSPPLATKERISPGIRGTHKLPRHRPLNRTQSAPLPQS  
TLAQLVIOQQHQQFLEKQKQYQQQIHMNKLKSKSIEQLKQPGSHLFEAEELQGDQAMQ  
EDRAPSSGNSTRSDSSACVDDTLGQVGAVKVKEEPVDSDEDAQIQEMESGEQA AFMQQV  
IGKDLAPGFVIKVII

FIG. 2E





23/173

HDAC9a	273	SGPSSPNNGPTG	SVTENETSVLP	PTPHAEOMV	SOORILLIHED	SMNLLSLYT	SPSLPNITL
HDAC9	273	SGPSSPNNGPTG	SVTENETSVLP	PTPHAEOMV	SOORILLIHED	SMNLLSLYT	SPSLPNITL
HDAC4	298	SGPSSPNNGPTG	SVTENETSVLP	PTPHAEOMV	SOORILLIHED	SMNLLSLYT	SPSLPNITL
HDAC4	298	SGPSSPNNGPTG	SVTENETSVLP	PTPHAEOMV	SOORILLIHED	SMNLLSLYT	SPSLPNITL
HDAC9a	333	GLPAVPSOLNAS	NSLKEKOKCET	OTLROGVPL	PGYGGSI	PASSSHPHVT	LEGKPPNSSH
HDAC9	333	GLPAVPSOLNAS	NSLKEKOKCET	OTLROGVPL	PGYGGSI	PASSSHPHVT	LEGKPPNSSH
HDAC4	357	GLPATGPSAGT	AGQQ-DTER	LTLPALQQR	LSLFGTHL	IPYLSIS	SP--PIERDG--GA
HDAC4	357	GLPATGPSAGT	AGQQ-DTER	LTLPALQQR	LSLFGTHL	IPYLSIS	SP--PIERDG--GA
HDAC9a	393	OALLLOHLL	LKEOMROOK	LLVAGG	--VPLHPOS	PLATKERI	SPGIRGTHKLPRHRPLNRTO
HDAC9	393	OALLLOHLL	LKEOMROOK	LLVAGG	--VPLHPOS	PLATKERI	SPGIRGTHKLPRHRPLNRTO
HDAC4	411	SPLLLOHML	VLLLEQPPAQ	APLVTGL	GLALPLHAQS	--LVGADR	RVSP--SIHKLROHRPLGRTO
HDAC9a	451	SAPLPQ--	STLAQ	LVIOOOH	QOFLEKOKO	--YOOOI	HMNKLLSKSIEOLKOPGSHLEAE
HDAC9	451	SAPLPQ--	STLAQ	LVIOOOH	QOFLEKOKO	--YOOOI	HMNKLLSKSIEOLKOPGSHLEAE
HDAC4	467	SAPLPQNAQ	ALQHLVIOOOH	QOFLEKOKO	QFQOOQ	LQMNKI	IPKPSERPARQFESHPEETE
HDAC9a	507	EELQGD	OAMOE	DRAPSSGNSTR	--SDSSA	CVDDTLGOV	GAVKVEEPVDSDEDAOIOEMES
HDAC9	507	EELQGD	OAMOE	DRAPSSGNSTR	--SDSSA	CVDDTLGOV	GAVKVEEPVDSDEDAOIOEMES
HDAC4	527	EELREH	QALLLE	QORIHOL	RNYQASME	AGIPVSFG	CHRFISR
HDAC4	527	EELREH	QALLLE	QORIHOL	RNYQASME	AGIPVSFG	CHRFISR
HDAC9a	566	GEOAAFM	QOQVIGK	DLAPGF	VIKVI	I-----	-----
HDAC9	566	GEOAAFM	QOQVIGK	DLAPGF	VIKVI	I-----	-----
HDAC4	587	ELLFRQ	QALLLE	QORIHOL	RNYQASME	AGIPVSFG	CHRFISR
HDAC4	587	ELLFRQ	QALLLE	QORIHOL	RNYQASME	AGIPVSFG	CHRFISR

FIG. 3B

24/173

HDAC9a	626	PLQPGSATGIAYDPLMLKHOCVCGNSTTHPEHAGRIOSIWSRLOETGLLNKCERI	OGRKA
HDAC9	626	PLQPGSATGIAYDPLMLKHOCVCGNSTTHPEHAGRIOSIWSRLOETGLLNKCERI	OGRKA
HDAC4	647	ETKPERFTIGLVYDTLMLKHOCVCGNSTTHPEHAGRIOSIWSRLOETGLLNKCERI	OGRKA
HDAC9a	686	SLEETOLVHSEHSLLYGTNPLDGOKLDPRILLGDDSOKEFFSSLPCCGLGVDSDTI	WNEI
HDAC9	686	SLEETOLVHSEHSLLYGTNPLDGOKLDPRILLGDDSOKEFFSSLPCCGLGVDSDTI	WNEI
HDAC4	707	TLLEETOLVHSEHSLLYGTNPLNRKLDSEKLLGSLASVVR-LPCGGVGVDS	TIWNEI
HDAC9a	746	HSSGAARMVAVGCVIELASKVASGELKNGFAVVRPPGHAAEESTAMGFCFFNSVAIT	TAKYL
HDAC9	746	HSSGAARMVAVGCVIELASKVASGELKNGFAVVRPPGHAAEESTAMGFCFFNSVAIT	TAKYL
HDAC4	766	HSAAGARLAVGCVIELVFKVATIGELKNGFAVVRPPGHAAEESTIPMGFCYFENSVA	VAAKIL
HDAC9a	806	RDOLNISKILLVDLVHHNGTQOAFYADPSILYISLHRYDEGNFFPGSGAPNEV	RFISL
HDAC9	806	RDOLNISKILLVDLVHHNGTQOAFYADPSILYISLHRYDEGNFFPGSGAPNEV	GTGLG
HDAC4	826	QORLSVSKILLVDLVHHNGTQOAFYSDPSVLVMSLHRYDDGNFFPGSGAPDE	VGTPG
HDAC9a	866	EPHFYLYLSGNCITIA	
HDAC9	866	EGYNINIAWTGGLDPPMGDVEYIEAFRTIVKFAKEFDPMVLVSA	GFDALLEGHTPLGG
HDAC4	886	VGENVNMATIGGGLDPPMGDAEYLAAFRTIVVMEHASEFAFDVVLVSSG	FDAVEGHPTPLGG
HDAC9a	926	YKVTAKCFCHLTQOLMTLADGRVVLALLEGCHDLTAICDASEACV	NALLGNELLEPIAEDIL
HDAC9	946	YNLSFARCFGYLTQOLMGLACGRIVLALLEGCHDLTAICDASEACV	SALLGNELDPIPEKVL
HDAC9a	986	HQSPNMNAVISLQKIIEIOSMSLKFS	
HDAC9	1006	QORPNANAVRSMKVMIEIHSKYWRCLQRTTSTAGRSLIEAOTCENEAE	TVTAMASLSVG
HDAC9a			
HDAC9			
HDAC4			
HDAC9a	1066	VKPAEKRPDEEPMEEPPPL	

FIG. 3C

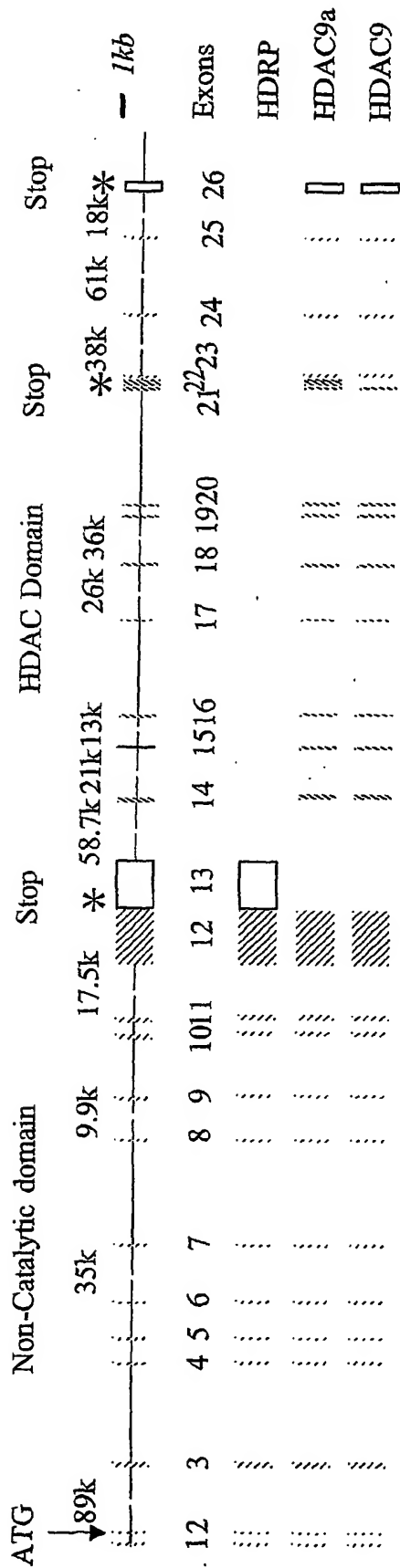


FIG. 4

25/173

FIG. 5A
FIG. 5B
FIG. 5C
FIG. 5D

FIG. 5

26/173

1 /<sup>1</sup>ggggaagaga ggcacagaca cagataggag aagggcacog gctggagcca ctbgcaggac tgagggtttt tgcaacaaaa  
ccctagcagc ctgaagaact

101 ctaagccag/<sup>2</sup>a tggggtggct ggacgagagc agctcttggc tcagcaaaaga ATGCACAGTA TGATCAGCTC AGT/<sup>3</sup>GGATGTG  
AAGTCAGAAG TTCCGTGTGGG

201 CCTGGAGCCC ATCTACACCTT TAGACCTAAG GACAGACCTC AGGATGATGA TGCCCGTGGT GGACCCCTGTT GTCCGTGAGA  
AGCAATTGCA GCAGGAATTA

301 CTTCTTATCC AGCAGCAGCA ACAATCCAG AAGCAGCTTC TGATAGCAGA GTTTCAGAAA CAGCATGAGA ACTTGACACG  
GCAGCACCAG GCTCAGCTTC

401 AGGAGCATAT CAAG/<sup>4</sup>GAACTT CTAGCCATAA AACAGCAACA AGAATCCTA GAAAAGGAGC AGAACTGGA GCAGCAGAGG  
CAAGAACAGG AAGTAGAGAG

501 GCATCGCAGA GAACAGCAGC TTCCTCTCTT CAGAGGCCAA GATAGAGGAC GAGAAAG /<sup>5</sup>GGC AGTGGCAAGT ACAGAAGTAA  
AGCAGAAGCT TCAAGAGTTC

601 CTA CTACTGAGTA AATCAGCAAC GAAAGACACT CCAACTAATG GAAAAAATCA TTCCGTGAGC CGCCATCCCA AGCTCTGGTA  
CACG/<sup>6</sup>GCTGCC CACCACAT

701 CATTTGATCA AAGCTCTCCA CCCCTTAGTG GAACATCTCC ATCCTACAAG TACACATTAC CAGGAGCACA AGATGCAAAG  
GATGATTTC CCCTTCGAAA

FIG. 5A

27/173

801 AACT/GCCTCT GAGCCCACT TGAAGTGG GTCCAGTTA AACAGAAAG TGGCAGAG GAGAAGCAGC CCCTTACTCA  
GGCGGAAGGA TGGAAATGTT  
 901 GTCACITCAT TCAAGAAGCG AATGTTTGAG GTGACAG /<sup>8</sup>AAT CCTCAGTCAG TAGCAGTTCT CCAGGCTCTG GTCCAGTTTC  
 ACCAAACAAT GGGCCAACTG  
 1001 GAAGTGTAC TGAAATGAG ACTTCGGTTT TGGCCCCCTAC CCTTCATGCC GAG /<sup>9</sup>CAATGG TTTCACAGCA ACGCATTTCTA  
 ATTCAAGAAG ATTCCATGAA  
 1101 CCCTGCTAAGT CTTTATACCT CTCCTTCTTT GCCCAACATT ACCTTGGGGC TTCCCGGCAGT GCCATCCCAG CTCAATG /<sup>10</sup>CTT  
 CGAATTCACCT CAAAGAAAAG  
 1201 CAGAAGTGTG AGAGCGAGAC GCTTAGGCAA GGTGTTCTTC TGCCTGGGCA GTATGGAGGC AGCATCCCGG CATCTTCCAG  
 CCACCCITCAT GTTACTTTAG  
 1301 AGGGAAGCC ACCCAACAGC AGCCACCAGG CTCTCCTTGA GCATTATTAT TTGAAGAAGC AAATGGGACA GCAAAAGCTT  
 CTTGTAGCTG/<sup>11</sup> GTGGAGTTCC  
 1401 CTTACATCCT CAGTCTCCCT TGGCAACAAA AGAGAGAAAT TCACCTGGCA TTAGAGGTAC CCACAAATG CCCCCTCACA  
 GACCCCTGAA CCGAACCAG  
 1501 TCTGCACCTT TGCCTCAGAG CAGTTGGCT CAGTGGTCA TTCAACAGCA ACACCAGCAA TTTCTTGGAGA AGCAGAAGCA  
 ATACCAGCAG CAGATCCACA  
 1601 TGAACAAA/<sup>12</sup>CT GCTTTGAAA TCTATTGAAC AACTGAAGCA ACCAGGCAGT CACCTTGAGG AAGCAGAGGA AGAGCTTCAG  
 GGGGACCAGG CGATGCAGGA

FIG. 5B

28/173

1701 AGACAGAGCG CCCTCTAGTG GCAACAGCAC TAGGAGCGAC AGCAGTGCTT GTGTGGATGA CACACTGGGA CAAGTTGGGG  
CTGTGAAGGT CAAGGAGGAA

1801 CCAGTGGACA GTGATGAAGA TGCTCAGATC CAGGAATGG AATCTGGGA GCAGGCTGCT TTTATGCAAC AG  
/13GTAATAGG CAAAGATTIA GCTCCAGGAT TTGTAATTAA AGTCATTATC TGA..... /14CCTTTCCT GGAACCCAGG CACACACGTG

1901 CGCTCTCTGT GCGCCAAAGCT CCGCTGGCTG CCGTTGGCAT GGATGGATTA GAGAAACACC GTCTCGTCTC CAGGACTCAC  
TCTTCCCCTG CTGCCCTCTGT

2001 TTACCTCAC CCAGCAATGG ACCGCCCCCT CCAGCCTGGC TCIGCAACTG /15GAATTGCTTA TGACCCCTTG ATGCTGAAAC  
ACCAGTGGCT TTGTGGCAAT

2101 TCCACCACCC ACCCTGAGCA TGCTGGAGCA ATACAGAGTA TCCTGTCAGG ACTGCAAGAA ACTGGGCTGC TAAATAAATG  
TGAG/16CGAATT CAAGTCGAA

2201 AAGCCAGCCT GGAGGAAATA CAGCTTGTTT CAGCTTGTTT CTAATTACCT TGTTGGTGGAC TTGGG/18GTGGA CAGTGACACC ATTTGGAAATG  
AAGCTGGACC CCAGGATACT

2301 CCTAG/17GTGAT GACTCTCAA AGTTTCTTTC CTCAATTACCT TGTTGGTGGAC TTGGG/18GTGGA CAGTGACACC ATTTGGAAATG  
AGCTACACTC GTCCGGTGCT

2401 GCAGGCATGG CTGTGGCTG TGTATCGAG CTGGCTTCCA AAGTGGCTC AGGAGAGCTG AAGA /19ATGGGT TTGCTGTGTGT  
GAGGCCCCCT GGCCATCAG

2501 CTGAAGAATC CACAGCCATG /20GGGTCTGCT TTTTAAATC AGTTGCAATT ACCGCCAAAT ACTTGAGAGA CCAACTAAAT  
ATAAGCAAGA TATTGATTGT

FIG. 5C

29/173

2601 AGATCTG/<sup>21</sup>GAT GTTCACCATG GAAACGGTAC CCAGCAGGCC TTTTATGCTG ACCCAGCAT CCTGTACATT TCACTCATC  
GCTATGATGA AGGGAACTTT

2701 TTCCCTGGCA GTGGAGCCCC AAATGAGG/<sup>22</sup>TT CGGTTATTT CTTTACAGCC CCACTTTAT TTGTATCTTT CAGGTAATTG  
CATTGCATGA ttacccctaa

2801 ttttcttctc ctttgctggt gttttaatt acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa  
gtgcataacc cagagcactg

2901 tttgtcaggg aaggttgggc tgatttgatg tgttgttga tgtttattc aagagctccc atgtgcttgt tttcctctct  
tcttcttctc ttccatttgc

3001 tctcttctct gccaccgtg gtgtgtcttt ctcttcccag/<sup>23</sup>gttggaaacag gccttggaga aggtacaat ataaatatg  
cctggacagg tggccttgat

3101 cctcccatgg gagatgttga gtaccttgaa gcattcag/<sup>24</sup>ga ccacgtgtaa gcctgtggcc aaagagtgtg atccagacat  
ggtcttagta tctgctggat

3201 ttgatgcatt ggaaggccac accctctctc taggagggtta caaagtacg gcaaaatg/<sup>25</sup>tt ttggtcattt gacgaagcaa  
ttgatgacat tggctgatgg

3301 acgtgtggtg ttggctctag aaggaggaca tgatctcaca gccatctgtg atgcatcaga agcctgtgta aatgcccttc  
taggaaatga g/<sup>26</sup>ctggagcca

3401 ctgcagaag atattctcca ccaaagcccg aatatgaatg ctgttatttc ttacagaag atcattgaaa ttcaaagtat  
gtctttaag ttctcttaa....

FIG. 5D



30/173

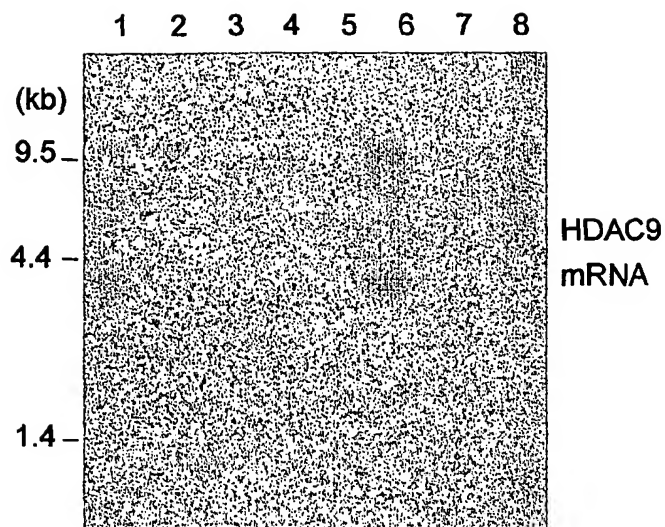


FIG. 6A

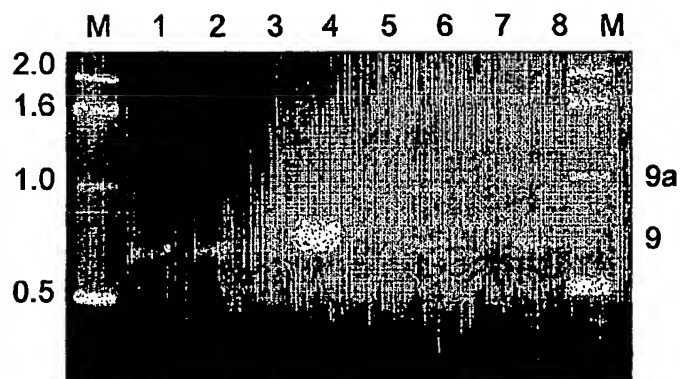


FIG. 6B

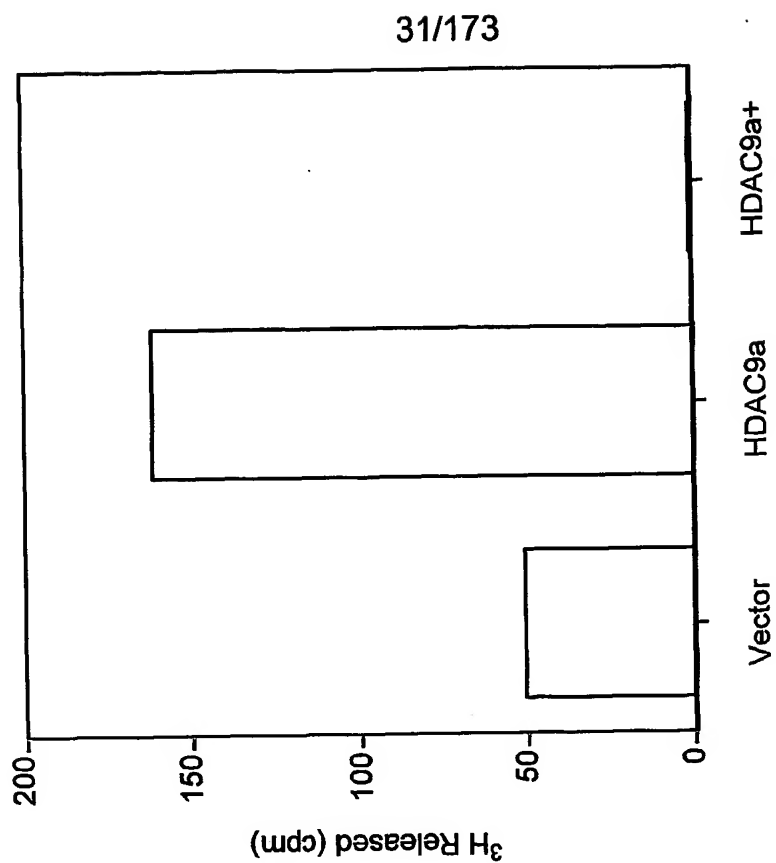


FIG. 8

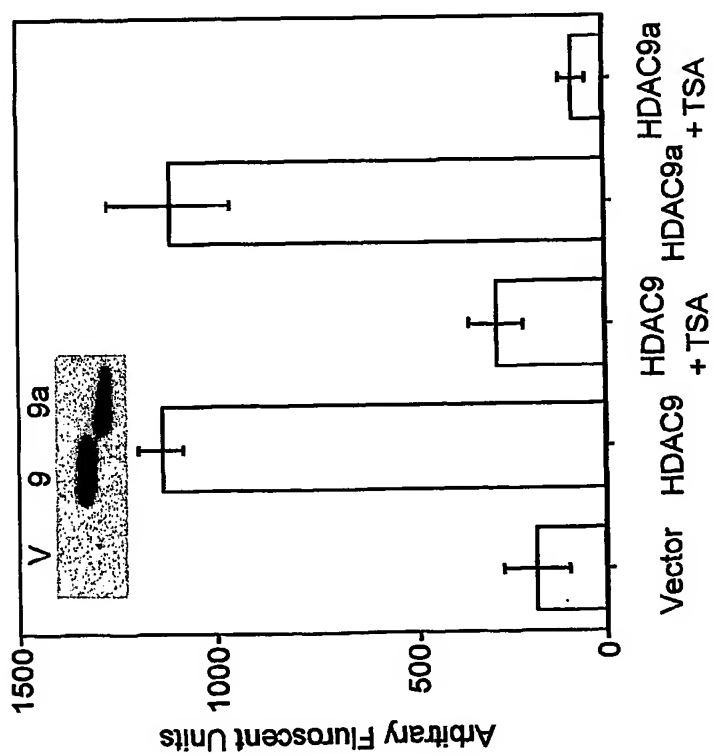


FIG. 7

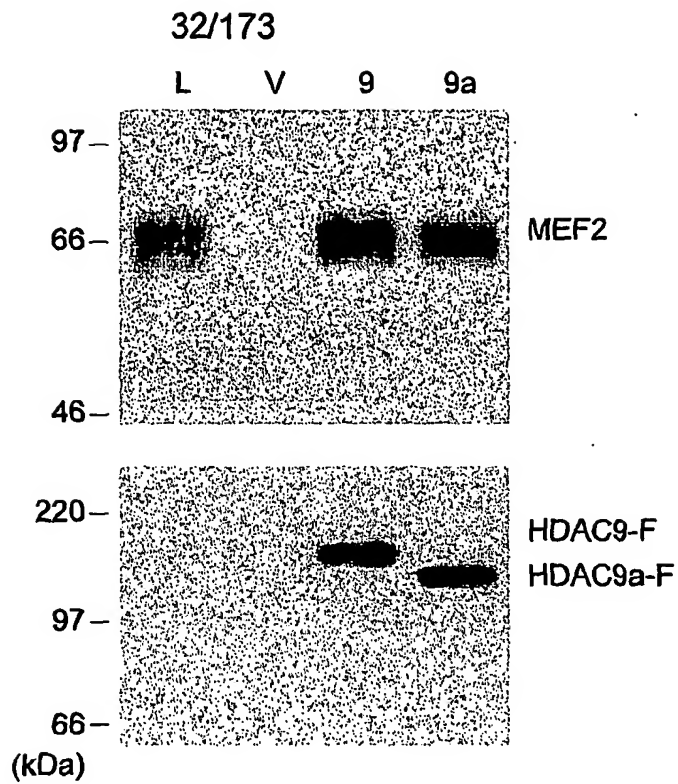


FIG. 9A

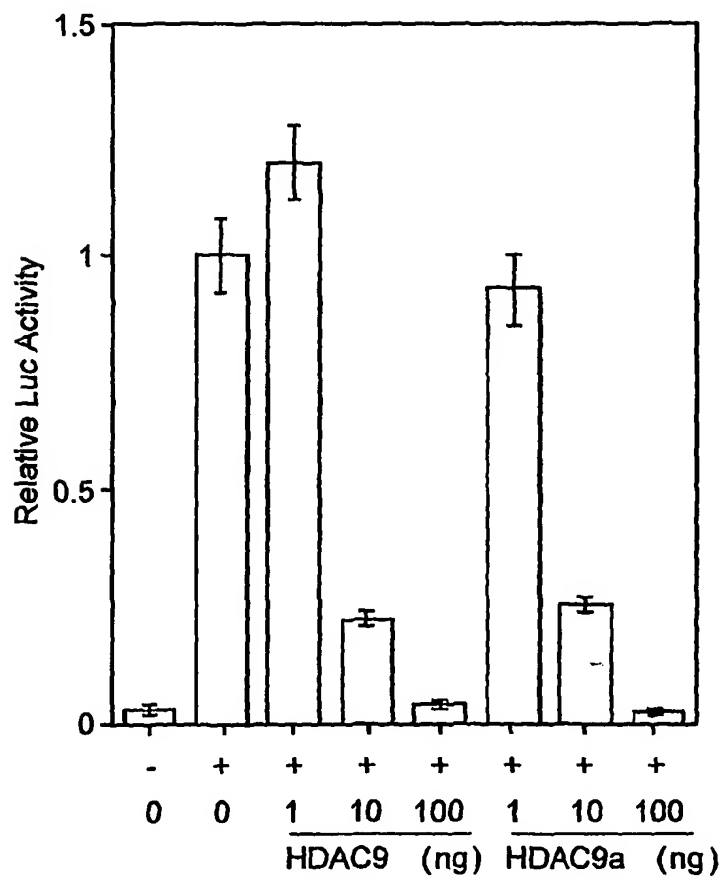


FIG. 9B

33/173

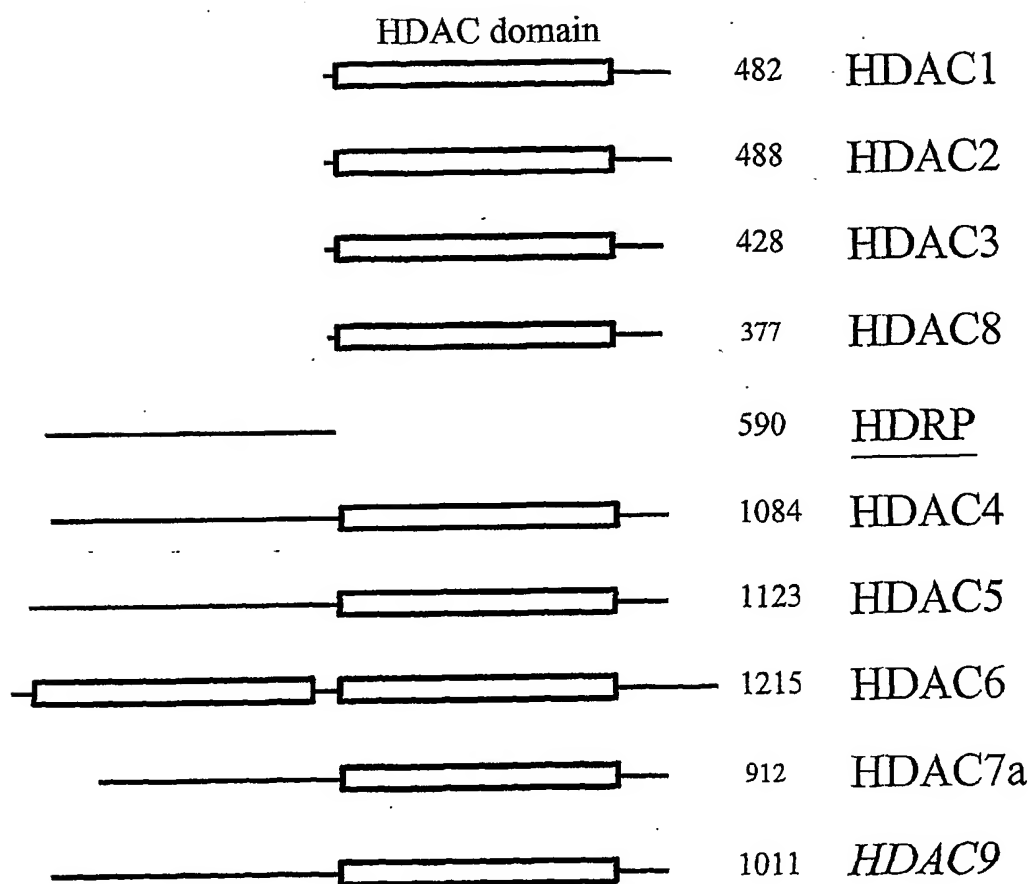


FIG. 10

34/173

FIG. 11A
FIG. 11B
FIG. 11C
FIG. 11D
FIG. 11E
FIG. 11F

FIG. 11

FIG. 11A

cccattccattcaggctgcgcaactgttgggaaggcgatcggcgccctctcgtattaccgagctggcgaaaggg  
ggatgtgctgcaaggcgattaagtgggtaacgccaggggtttccagtcacgacgttgtaaaacgacggccagtgccaagct  
gatctaataatattggccattagccatattattcattggtatatagcataaataattggcctattggccattgcatacgttgatcca  
tatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagtatttaataagtaataattacg  
gggtcattagttcatagcccatatatggagttccggttacataacttacggtaaatggcccgctggcgaccgccagcgaccc  
ccggccgttgacgtcaatagtagctatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacg

gtaactcccacttggcagtagatcaagtgtatcatatgccaaagtcgccccctattgacgtcaatgacggtaaatggcccgcct  
 agcattatgccagtagacattacggaggttctacttggcagtagacatctactacgtattagtcacgtattaccatggtgatgcg  
 gtttggcagtagaccaatggcgtgtagcgggttgactacgggattccaagtcctcaccattgacgtcaatgggaggtt  
 tgtttggcaccaaaatcaacgggacttccaaaatgtcgtataacccccggcgttgacgcaaatggcggtagcggtgtacg  
 gtggaggtctatataagcagagctcgttttagtgaaccgtcagaattcaagcttgcggccgcagatctatcgatctgcaggatatac  
 (EcoRV)  
*acc*

35/173

ATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAAGTTCTCTGTGGG  
 CCTGGAGCCCATCTCACCTTTAGACCTAAGGACAGACCTCAGGATGATGA  
 TGCCCGTGTGGACCCCTGTTGTCCGTGAGAAAGCAATTGCAGCAGGAATTA  
 CTTCTTATCCAGCAGCAGCAACAAATCCAGAAAGCAGCTTCTGTATAGCAGA  
 GTTTCAGAAACAGCATGAGAACTTGACACGGCAGCACCCAGGCTCAGCTTC  
 AGGAGCATATCAAGGAACCTTAGCCATAAAACAGCAACAAAGAACTCCTA  
 GAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAAACAGGAAGTAGAGAG  
 GCATCGCAGAGAACAGCAGCTTCTCTCTCAGAGGCAAGATAGAGGAC  
 GAGAAAGGCAGTGGCAAGTACAGAAAGTAAAGCAGAAAGCTTCAAGAGTTC  
 CTACTGAGTAAATCAGCAACGAAAGACACTCCAATAATGGAAAAAATCA  
 TTCCGTGAGCCGCATCCCAGCTCTGGTACACGGCTGCCACACACAT  
 CATTGGATCAAAGCTCTCCACCCCTTAGTGGAACATCTCCATCTCTACAAG

FIG. 11B

36/173

TACACATTACCAGGAGCACAAAGATGCAAAGGATGATTTCCCCCTTCGAAA  
AACTGCCCTCTGAGCCCAAATTGAAGTGGTCCAGGTTAAACAGAAAG  
TGGCAGAGAGGAGAACGCCCTTACTCAGGCGAAGGATGGAATGTT  
GTCACCTTCATTCAAGAAGCGAATGTTTGAGGTACAGAATCCTCAGTCAG  
TAGCAGTTCTCCAGGCTCTGGTCCCAGTTCACCAACAATGGGCCAACTG  
GAAGTGTACTGAAAATGAGACTTCGGTTTGGCCCCCTACCCCTCATGCC  
GAGCAAAATGGTTTCACAGCAACGCATTCTAATTCATGAAGATTCATGAA  
CCTGCTAAGTCTTTATACCTCTCTCTTTTGCCCAACATTACCTTGGGC  
TTCCCCGAGTGCCATCCCAGCTCAATGCTTCGAATTCACCTCAAGAAAAG  
CAGAAAGTGTAGACGCAGACGCTTAGGCAAGGTGTTCTCTGCTGGGCA  
GTATGGAGGAGCATCCCGGCATCTTCCAGCCACCTCATGTACTTTAG  
AGGAAAGCCACCCAAACAGCAGCCACCAGGCTCTCTGTCAGCATTTATTA  
TTGAAAGAAACAAATGCGACAGCAAAAGCTTCTTGAGCTGGTGGAGTTCC  
CTTACATCCTCAGTCTCCCTTGGCAACAAAGAGAGAAATTCACCTGGCA  
TTAGAGGTACCCACAAATTGCCCCGTACAGACCCCTGAACCGAACCCAG  
TCTGCACCTTTGCCCTCAGAGCACGTTGGCTCAGCTGGTCATTCAACAGCA  
ACACCAGCAATTCTTGGAGAGCAGAGAAGCAATACCAGCAGCAGATCCACA  
TGAACAAAACCTGCTTTCGAAATCTATTGAACAACTGAAGCAACCAGGCAGT  
CACCTTGAGGAAGCAGAGGAAGAGCTTCAGGGGGACCAGGCGATGCAGGA  
AGACAGAGCGCCCTCTAGTGGCAACAGCACTAGGAGCGACAGCAGTGCTT  
GTGTGGATGACACACTGGGACAAAGTTGGGGCTGTGAAGTCAAGGAGGAA  
CCAGTGGACAGTGATGAAGATGCTCAGATCCAGGAAATGGAATCTGGGGA  
GCAGGCTGCTTTTATGCAACAGCCTTTCCTGGAAACCCACGCACACACGTG  
CGCTCTCTGTGCGCCAAGCTCCGCTGGCTGCGGTTGGCATGGATGGATTA

FIG. 11C

37/173

GAGAAACACCGTCTCGTCTCCAGGACTCACTCTTCCCCTGCTGCCCTCTGT  
TTTACCTACCCAGCAATGGACCGCCCTCCAGCCTGGCTCTGCAACTG  
GAATTGCCCTATGACCCCTTGATGCTGAAACACCAAGTCCGTTTGTGGCAAT  
TCCACCACCCCTGAGCATGCTGGACGAATACAGATATCTGGTCACG  
ACTGCAAGAAACTGGGCTGCTAAATAAATGTGAGCGAATTCAAGTCTGAA  
AAGCCAGCTGGAGGAAATACAGCTTGTTCATTCTGAACATCACTCACTG  
TTGTATGGCACCACCCCTGGACGGACAGAAAGCTGGACCCAGGATACT  
CCTAGGTGATGACTCTCAAAAAGTTTTTTTCCCTCATTA CCTTGTGTGGAC  
TTGGGGTGGACAGTGACACCATTTTGGAAATGAGCTACACTCGTCCGGTGCT  
GCACGCATGGCTGTTGGCTGTGTCACTCGAGCTGGCTTCCAAAAGTGGCCTC  
AGGAGAGCTGAAGAAATGGGTTTGTCTGTTGTGAGGCCCTTGGCCCATCACG  
CTGAAGAATCCACAGCCATGGGGTTCTGCTTTTAAATTCAGTTGCAATT  
ACCGCCAAATACTTGAGAGACCAACTAAATAAAGCAAGATAATTGATTGT  
AGATCTGATGTTTCAACCATGGAAACGGTACCCAGCAGGCCCTTTTATGCTG  
ACCCAGCATCCGTGTACATTTCACTCCATCGCTATGATGAAGGAACTTT  
TTCCCTGGCAGTGGAGCCCAAAATGAGGTTGGAAACAGGCCCTTGGAGAAAG  
GTACAATATAAATATTGCCCTGGACAGGTGGCCTTGATCCTCCCATGGGAG  
ATGTTGAGTACCTTGAAGCATTCAGGACcaTCGTGAAGCCTGTGGCCAAA  
GAGTTTGATCCAGACATGGTCTTAGTATCTGCTGGAATTGATGCATTGGA  
AGGCCACACCCCTCCTTAGGAGGGTACAAAGTGACGGCAAAATGTTTGTG  
GTCAATTGACGAAGCAATTGATGACATTGGCTGATGGACGTGTGGTGTG  
GCTCTAGAAGGAGGACATGATCTCACAGCCATCTGTGATGCATCAGAAGC  
CTGTGTAAATGCCCTTCTAGGAAATGAGCTGGAGCCACTTGCAGAAAGATA  
TTCTCCACCAAAGCCCGAATATGAATGCTGTTATTCTTTTACAGAAAGATC  
ATTGAAATTCAAAGTATGTCTTTTAAAGTTCTCT

FIG. 11D





[illegible]

40/173

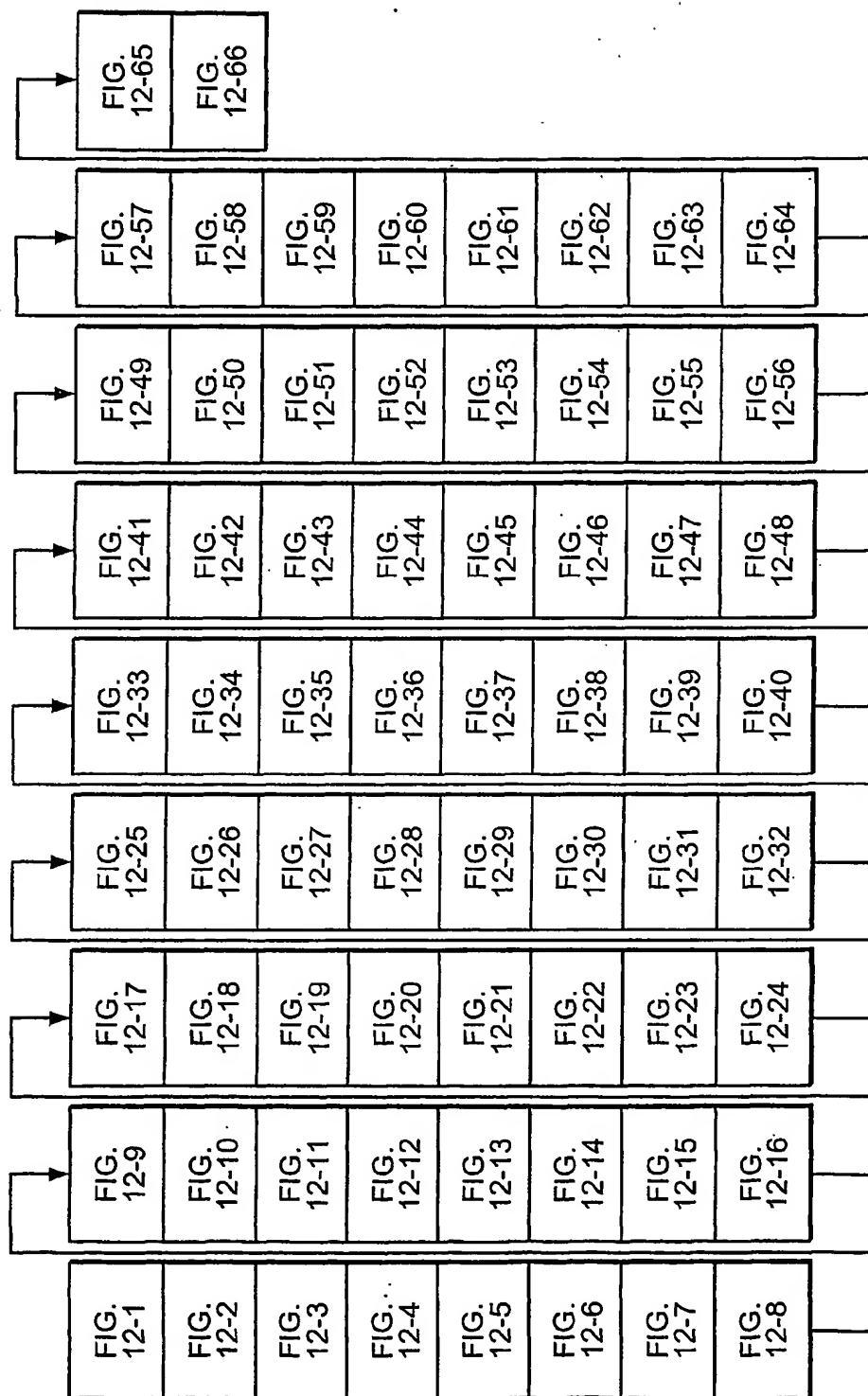


FIG. 12

pFLAG-CMV-5b-HDAC9

7699 base pairs

Graphic map | Table by enzyme name

BglI	FspI	AviII	BstMCI	EarI	MspAII
1 to 75	base pairs	cccattcgccattcaggctgcgcaactgttgggaaggcgatcggcgggcctcttcgctattacgccagctgg	PvuI BsiEI	Eam1104I	PvuII
41/173	base pairs	gggtaagcggtaagtccgacgcgttgacaacccttcccgctagccacgcccggagaagcgataatgcggtcgacc	BsaOI	Ksp632I	NspBII
76 to 150	base pairs	cgaaagggggatgtgctgcaaggcgattaaagtgggtaacgccagggttttcccagtcacgacgttgtaaaacg	Acc16I	BspCI	Ple19I
76 to 150	base pairs	gctttcccctacacgacgttccgctaattcaaccattgcgggtcccaaaagggtcagtgctgcaaacattttgc	Bsh1285I	Ksp632I	NspBII

FIG. 12-1

42/173

EaeI	MscI	
	CfrI	
SspI MluNI		
acggccagtgccaagctgatctaataatcaataattggccattagcccatattattcattggttatatagcataaatcaa		
base pairs		
tgccggtcacgggttcgactagattagttataaacgggtaatcgggtataataagtaaccaatatatcgtatttagtt		
151 to 225		
CfrI	EaeI	
	Bali	
MscI	SspBI	
MluNI	Bsp1407I	
SspI EaeI BsrDI		
tattggctattggccattgcatacgttggtatccatatcataaataatgtacattttatatattgggtcatgtccaacatt		
base pairs		
ataaccgataaacgggtaacgtatgcaacataggtatattatatacatgtaaataataaccgagtagcaggttgtaa		
226 to 300		
CfrI	BsrGI	
Bali		

FIG. 12-2

HincII	SpeI	VspI	
		PshBI	
accgccatgttgacattgattgactagttattagtaatacaattacggggtcattagttcatagcccata			
base pairs			
tggcggtaacaactgtaactaataactgatcaataattatcattagttaatgccccagtaatacaagtatcgggtat			
301 to 375			
HindII	AcI	AsnI	
	NI	AI	
tatggagttccgcgttacataacttacggtaaatggccccgcctggcgaccgccagccccgcgcttgacg			
base pairs			
atacctcaaggcgcaatgtattgaatgccatttacggggcgaccgctggcgggtcgctggggcgggcaactgc			
376 to 450			

43/173

HincII	BstMCI	
AcyI	BsaOI	
HincII	BglI	
tatggagttccgcgttacataacttacggtaaatggccccgcctggcgaccgccagccccgcgcttgacg		
base pairs		
atacctcaaggcgcaatgtattgaatgccatttacggggcgaccgctggcgggtcgctggggcgggcaactgc		
376 to 450		
HindII	Bsh1285I	
	BsiEI	

FIG. 12-3

AatII  
 BbiII  
 tcaatagtgacgtatgttcccatagtaacgccaatagggaacttccattgacgtcaatgggtggagtatttacgg  
 base pairs  
 agttatcactgcatacaagggtatcattgcgggttatccctgaaaggtaactgcagttaccacctcataaatgcc  
 451 to 525  
 Hsp92I

BbiII  
 Hin1I  
 AcyI AatII

Msp17I  
 BsaHI  
 Hsp92I

44/173

BglI  
 NdeI  
 taaactgcccaacttggcagtagtcaatcaagtgatcatatgcccaagtcgccccctattgacgtcaatgacggtaaa  
 base pairs  
 atttgacgggtgaaccgtcatgttagttcacatagtagtatacgggttcaggcgggggataaactgcagttactgccattt  
 526 to 600

BbiII  
 Hin1I  
 AcyI AatII

Msp17I  
 BsaHI  
 Hsp92I

FauNDI

FIG. 12-4

BstSNI  
SnaBI

tgccccgcctagcattatgcccagtagaccttacgggagtttcctacttggcagtagacatctacgtattagtc  
base pairs  
accggcggtacgtaatacgggtcatgtactggaatgccctcaaaggatgaaccgtcatgtagatgcataatcag  
601 to 675

BsaAI  
Eco105I

45/173

NcoI Bsp19I  
StyI BstDSI  
EcoT14I

atcgctattaccatgggtgatgcggttttggcagtagacaccaatgggcgtggatagcgggtttgactcacggggattt  
base pairs  
tagcgataatggtagccactacgcccacaaaccgtcatgtggttaccgcacacctatcgccaaactgagtgccctaaa  
676 to 750

BssT1I  
ErhI Eco130I  
DsaI MslI

FIG. 12-5



BblII  
 HinII  
 AccB1I  
 AcyI AatII  
 BshNI  
 ccaagtctccacccattgacgtcaatgggagtttggcaccaaaatcaacgggactttccaaaatgtcgt  
 base pairs  
 ggttcagaggtgggtaactgcagttaccctcaaaacccgtggttttagttgccctgaaagggttttacagca  
 751 to 825  
 Msp17I  
 BsaHI  
 Hsp92I  
 BanI  
 Eco64I  
 46/173  
 HincII  
 BanII  
 Eco24I  
 EcoICRI  
 aataaccccgcccggttgacgcaaatgggcggtaggcgtgtacggtgggaggtctataaagcagagctcgttta  
 base pairs  
 ttattgggcggggcaactgcggtttaccgcgccatccgcacatgccaccctccagatatattcgtctcgagcaaat  
 826 to 900  
 HindII  
 Ecl136II  
 SacI

FIG. 12-6

47/173

```
FriOI
Ssti
BsiHKAi
Bbv12i
AspHI
gtgaaccgtcagaattcaagcttgccgcagatctatcgatctgcaggatatcaccatgcacagtatgatcag
base pairs
cacttggcagtccttaagtgcgaacgccggcgtctagatagctagacgtcctctatagtggtacgtgtcatactagtc
901 to 975
Psp124BI EcoRI
EaeI Eco52I BglII BscI BspXI BstSFI FbaI
CfrI EclXI BsiEI BseCI Bsu15I EcoRV
NotI BsaOI XhoII ClaI Bsp106I

Alw21I

CvnlI CvnI
AocI AocI
Bsu36I Bsu36I

ctcagtggtatgtgaagtcagaattcctgtggcctggagcccatctcacctttagacctaaaggacagacctcag
base pairs
gagtcacctacacttcagtccttcaaggacacccggacctcgggtagagtggaatatctggattcctgtctggagtc
976 to 1050

Eco81I Eco81I
Bse21I Bse21I
GsuI BanII
```

FIG. 12-7

gatgatgagcccggtggtggaccctgtgtggtccgtgagaagcaattgcagcaggaattacttcttatccagcagca  
base pairs  
ctactactacgggacccacacctgggacaaacaggcactcttcggttaacgtcgtccttaatgaagaatagggtcgtcgt  
1051 to 1125

DsaI                      DrdI                      MfeI                      Asp700I

BstDSI                      MunI                      XmnI

48/173

FIG. 12-8

AlwNI  
gcaacaaatccagaagcagcttctgatagcagagtttcagaaacagcatgagaacttgacacggcagcaccaggc  
base pairs  
cggtggttaggtcttcgtcgaagactatcgctcctcaagtctttgtcgtaactcttgaaactgtgccgtcggtggtccg  
1126 to 1200

49/173

BlpI	Eco57I	EcoNI	AlwNI
CellI			
tcagcttcaggagcatatcaaggaaacttctagccataaaacagcaacaagaactcctagaaaaggagcagaaact			
base pairs			
agtcgaagtcctcgatagttccttgaagatcggtattttgtcgttgttcttgaggatcttttcctcgctcttga			
1201 to 1275			
Bsp1720I			
Bpu1102I			

FIG. 12-9

BpmI  
ggagcagcagaggcaagaacaggaagtagagaggcatcgagagaaacagcagcttcctcctctcagaggcaaga  
base pairs  
cctcgtcgtctccggtcttgctccttcattctcctcgtagcgtctctgctcgaaggaggagagtctccgtttct  
1276 to 1350  
GsuI

50/173

EcoNI

HindIII  
tagaggacgagaaagggcagtggaagtagcagaagtaaacag aagcttcaagagttcctactgagtaaatcagc  
base pairs  
atctcctgctctttcccggtcacccgttcattgtcttcatttcgctc ttcgaagttctcaaggatgactcatttagtcg  
1351 to 1425

FIG. 12-10

Van91I	Van91I	
AccB7I	AccB7I	
aacgaaagacactccaactaatggaaaaaatcattccgtgagccgccatcccaagctctggtacacggctgcca		
base pairs		
ttgctttctgtgaggttgattaccttttttagtaaggcactcggcggtaggggttcgagaccatgtgccgacgggt		
1426 to 1500		
Esp1396I	Esp1396I	51/173
Pf1MI	Pf1MI	
ccacacatcattggatcaaaagctctccaccccttagtggaacatctccatccctacaagtacacattaccaggagc		
base pairs		
ggtgtgtagtaacctagtttcgagaggtggggaatcaccttgtagaggtaggatgttcatgtgtaatggtcctcg		
1501 to 1575		

FIG. 12-11

52/173

Alw21I	BstBI		
AspHI	Bpu14I	FriOI	
	Csp45I	Eco24I	
acaagatgcaaaaggatgattcccccttcgaaaaaactgcctctgagcccaacttgaaggcggtccagggttaa			
base pairs			
tggtctacgtttccctactaaaggggaagctttttgacggagactcgggttgaacttccacgccagggtccaattt			
1576 to 1650	SfuI Bsp119I	BanII	
BsiHKA I	NspV		
Bbv12I	LspI		

	BseRI	EcoNI
acagaaaagtggcagagaggagagaagcagcccccttactcaggcggaaggatggaaatgtgtcacttcattcaagaa		
base pairs		
tgtctttcaccggtctctcctcttcgtcgggggaatgagtcggccttcctaccctttacaacagtgaaagtaagtctt		
1651 to 1725		

FIG. 12-12

Van91I  
AccB7I  
BpmI PflMI

Van91I  
AccB7I

gcgaatgtttgaggtgacagaatcctcagtcagtagcagttctccaggctctggtcccagttcaccaacaatgg  
base pairs  
cgcttacaaaactccactgtcttaggagtcagtcacgtcaagaggtccgagaccagggtcaagtgggtttgtttacc  
1726 to 1800

GsuI  
Esp1396I  
AlwNI

Esp1396I  
PflMI

53/173

gccaaatggaagtgttactgaaaatgagacttcgggttttgccccctacccctcatgccgagcaaatggtttcaca  
base pairs  
cggttgaccttcacaatgacttttactctgaagccaaaacggggatggggagtagcggctcgtttaccaaagtgt  
1801 to 1875

FIG. 12-13



BsaMI  
Mva1269I  
BspMI  
XcmI  
gcaacgcattcttaattcatgaagattccatgaacctgctaagtctttatacctctccttcttggcccaacattac  
base pairs  
cgttgcgtaagattaagtacttctaaggtagtggacgattcagaaatatggagaggaagaaacgggttgtaatg  
1876 to 1950  
BsmI RcaI  
BspHI

54/173

ErhI  
BssT1I  
BstBI AcsI  
Bpu14I  
Csp45I  
Esp3I  
cttggggcttcccgcagtgccatcccagctcaatgcttc gaattcactcaaagaaagcagaagtgtgagacgca  
base pairs  
gaaccccgaaaggcggtcacggtagggtaggttacgaag ctttaagtgaagtcttcttcttcacactctgcgt  
1951 to 2025  
EcoT14I  
SfuI Bsp119I  
BsmBI  
StyI  
Eco130I  
NspV ApcI  
LspI EcoRI

FIG. 12-14

55/173

MsII  
gacgcttaggcaaggtgttcctctgctgctgggagtagtgaggcagcatccggcatcttccagccacctcatgt  
base pairs  
ctgcgaatccggtccacaaggagacggaccggtcatacctcgtagggccgtagaaggtcggtgggagtaca  
2026 to 2100

PstI  
SfiI  
tactttagagggaagccaccaaacagcagcaccaggctctc ctgcagcatttattgaaagaacaaatgcg  
base pairs  
atgaaatctccctttcgggtggtgtcgtcgggtggtccgagag gacgtcgtaataataactttcttgtttacgc  
2101 to 2175  
BstSFI

FIG. 12-15

56/173

<p>HindIII</p> <p>acagcaaaaagcttctttagctggtaggtcccttacatcctcagtcctcccttggcaacaaaagagagaatttc base pairs tgtcgttttcgaagaacatcgaccacctcaagggaatgtaggagtcagaggggaaccgttggtttctctctcttaaag 2176 to 2250</p>	<p>Eco130I</p> <p>StyI</p> <p>EcoT14I</p> <p>ApoI</p>
<p>Asp718I</p> <p>Acc65I</p> <p>BshNI</p> <p>acctggcattagaggtaccacacaaattgcccgcgtcacagacccctgaaccgaaccagtcctgcacctttgcctca base pairs tggaccgtaatctccatgggtgtttaacgggcagtgctctgggacttggcttgggtcagacgtggaaacggagtc 2251 to 2325</p>	<p>BssT1I</p> <p>ErhI</p> <p>AcSI</p> <p>BsgI</p>
<p>BanI</p> <p>KpnI</p> <p>AccB1I</p> <p>Eco64I</p>	

FIG. 12-16

57/173

Bpu1102I  
 Alw21I Bsp1720I  
 AspHI CelII  
 gagcacgttggctcagctggtcattcaacagcaacaccagcaattcttggagaagcagaagcaataaccagcagca  
 base pairs  
 ctcgTgcaaccgagtcgaccagtaagtgtcgTgtggtcgTtaagaacctcttcgTcttcgTtatggTcgTcgT  
 2326 to 2400  
 BsiHKA I PvuII  
 Bbv12I B1pI MspA1I  
 NspBII  
 BstBI  
 Bpu14I  
 Csp45I Eco57I  
 gatccacatgaacaaactgcttctcgaaatctattgaacaaactgaagcaaccaggcagtcaccttgaggaaagcaga  
 base pairs  
 ctaggTgtacttgtttgacgaaagcttttagataaacttgttgacttcgTtggtccgTcagTggaaactcccttcgTct  
 2401 to 2475  
 BstYI  
 BstX2I  
 SfuI Bsp119I  
 NspV  
 LspI

FIG. 12-17

58/173

EarI  
 Eam1104I  
 Asp700I  
 Bbv16II  
 BbsI  
 Bsp143II  
 ggaagagcttcaggggaccaggcgatgcaggaagacagagcgccctctagtggaacagcactaggagcgacag  
 base pairs  
 ccttcgaagtcctccctgggtccgctacgtccttctgtctcggggagatcacctgtgtcgtgatcctcgtgtgc  
 2476 to 2550  
 XmnI  
 Eco57I  
 Ksp632I  
 SspI  
 BpiI  
 HaeII  
 BpuAI  
 BstH2I

BcgI  
 cagtgccttggtggtgacacactgggacaagtggggctgtgaagggtcaaggagggaaccagtggaacagtgatga  
 base pairs  
 gtcacgaacacacactactgtgtgacccctgttcaaccccgacacttcagttcctccttggtcacctgtcactact  
 2551 to 2625

FIG. 12-18

MflI Van91I  
XhoII AccB7I  
agatgctcagatccaggaaatggaatctgggagcaggctgcttttatgcaacagcctttccttggaaacccacgca  
base pairs  
tctacgagtcctaggtcctttaccttagacccctcgtccgacgaaatacgttgcgaaaggaccttgggtgcgt  
2626 to 2700

BstYI Esp1396I  
BstX2I PflMI

59/173

PmaCI  
PmlI  
AflIII

NspBII

Esp3I

cacacgtgcgctctctgtgcgccaagctccgctggctgcggttggcatggattagagaaacacccgtctcgt  
base pairs  
gtgtgcacgcgagagacacgcggttcgaggcaccgacgccaaccgtacctaatctcttgtggcagagca  
2701 to 2775

MslI Eco72I

MspAII

BsmBI

BsaAI  
BbrPI

FIG. 12-19

60/173

	EarI		BsrDI	BpmI
	Eam1104I			
ctccaggactcactcttccccctgctgcctctgttttacctcaccagcaatggaccgccccctccagcctggctc				
base pairs				
gaggtcctgagtgagaaggggacgacggagagacaaaaatggagtgggtcgttacctggcgggggaggtcggaccgag				
2776 to 2850				
GsuI	Ksp632I			GsuI

	XcmI
tgcaactggaattgcctatgaccccttgatgctgaaacaccagtgcggtttgtggcaattccaccaccacctga	
base pairs	
acgttgaccttaacggatactggggaactacgactttgtggtcacgcaaacacaccttaagtggtgggtgggact	
2851 to 2925	

FIG. 12-20

sphI  
 BbuI  
 gcgtgctggacgaatacacagagtatctggtcacgactgcaagaaactgggctgctaataaatgtgagc gaattca  
 base pairs  
 cgtagacactgcttatgtctcatagaccagtgctgacgttctttgacccgacgatttatttacactcg cttaaagt  
 2926 to 3000  
 PaeI  
 NspI  
 AcsI  
 ApsI  
 EcoRI

61/173

AggTcGaaaagccagcctggaggaaatacagcttggtcattctgaacatcactcactgttgatggcaccaaccc  
base pairs  
tccagcttttcggtcggaacctcctttatgtcgaacaagtaagacttgtagtgagtgacaacataccgtggttggg  
3001 to 3075

BpmI  
AccB1I  
BshNI  
BanI  
Eco64I  
GsuI

FIG. 12-21



62/173

ErhI  
StyI Eco130I  
ECOT14I  
BstXI AlwNI  
cctggacggacagaagctggacccaggataactcctagtgatgactctcaaaagttttttcctcattaccttg  
base pairs  
ggacctgcctgtcttcgacctggggtcctatgaggatccactactgagagttttcaaaaaaggagtaatggaac  
3076 to 3150

BstTII  
AvrII  
BlnI

BsaWI BsgI  
tggaggacttggggtggacagtgacaccatttgggaatgagctacactcgtccggtgctgcacgcacgtggtgttg  
base pairs  
accacctgaacccacacctgtcactgtggtaaaccttactcgtatgtgagcaggccacgacgtgcgtaccgacaacc  
3151 to 3225

FIG. 12-22

CvnI CfrI  
AocI DraII EaeI  
Bsu36I Eco57I  
ctgtgtcatcgagctggcttccaaagtggcctcaggagagctgaagaatgggtttgctgtgtgagggccccctgg  
base pairs  
gacacagtagctcgaccgaaggtttcacccggagtcctctcgacttcttaccctaaacgacacactccgggggacc  
3226 to 3300 Eco81I  
Bse21I EcoO109I

63/173

MscI ErhI Eco130I  
BssT1I BstXI  
Eco57I MslI DsaI  
ccatcacgctgaagaatccacagccatggggtctgtctttttaattcagttgcaattaccggccaaatacttgag  
base pairs  
ggtagtgcgacttcttaggtgtcggtaccccaagacgaaaaataagtcaacgttaatggcggtttatgaactc  
3301 to 3375  
MluNI  
Bali  
EcoT14I  
StyI BstDSI  
NcoI Bsp19I

FIG. 12-23

BstX2I NcoI Bsp19I Asp718I SseBI  
 BstYI StyI BstDSI AccB1I  
 XhoII EcoT14I BshNI StuI  
 BsaI Eco147I  
 agaccaactaaataaagcaagataattgattgtagatctggatgttcaccatggaaacgggtaccacagcaggcctt  
 base pairs  
 tctggttgatttatattcgttctataactaacatctagacacctacaagtgggtacaccttgccatgggtcgtcgcggaa  
 3376 to 3450  
 Eco31I BglII BstT1I BanI KpnI AatI  
 MflI ErhI Eco130I Eco64I Pme55I  
 DsaI Acc65I  
 64/173

SspBI  
 Bsp1407I MslI Asp700I  
 ttatgctgacccagcatcctgtacatttcactccatcgctatgatgaagggaacttttccctggcagtgaggc  
 base pairs  
 aatacgactgggtcgtaggacatgtaaagtgaggtagcgataacttcccttgaaaaaggacccgtcacctcg  
 3451 to 3525  
 BsrGI XmnI

FIG. 12-24

<p>                     FriOI                      Eco24I                      cccaatgaggttggaacaggccttggaagaagggtacaataataattgcctggacaggtggccttgatccctcc                      base pairs                      gggtttactccaaccttgtccggaacctcttcccatgttatattataacgggacctgtccaccggaactaggagg                      3526 to 3600                      BanII                 </p>	<p>                     SseBI ErhI                      Eco147I                      StuI BssTII                      SspI                 </p>
<p>                     AatI StyI                      Pme55I Eco130I                      EcoT14I                 </p>	<p>                     MscI                      MluNI                      AspI                 </p>
<p>                     NcoI Bsp19I                      StyI BstDSI                      EcoT14I                 </p>	<p>                     BsaMI                      Mva1269I                      EaeI                 </p>
<p>                     catgggagatgttgagtaccttgaagcattcaggaccatcgtgaagcctgtggccaaagagtttgatccagacat                      base pairs                      gtaccctctacaactcatggaacttcgtaagtcctggtagcacttcggacacccgggtttctcaaaactagggtctgta                      3601 to 3675                      BssTII                      DsaI                      ErhI Eco130I                 </p>	<p>                     AtsI                      CfrI                      Tth111I                      BalI                 </p>

65/173

FIG. 12-25

Mph1103I  
 EcoT22I  
 Ppu10I  
 EcoNI  
 ggtcttagtatctgctggatttgatgcattggaaggccacacccctcctctaggagggtacaaagtgacgggcaaa  
 base pairs  
 ccagaatcatagacgacctaactaacgttaaccttcgggtgtggggaggagatcctcccatgtttcactgccgttt  
 3676 to 3750  
 NsiI  
 BseRI  
 66/173  
 MfeI  
 AflIII  
 XbaI  
 atgttttgggtcatttgacgaagcaattgatgacattggctgatggacgtgtggtgttggtctctagaaggaggaca  
 base pairs  
 tacaaaaaccagtaaaactgcttcgttaactactgttaaccgactacctgcacaccacaacccgagatcttccctcctgt  
 3751 to 3825  
 MunI

FIG. 12-26

Mph1103I  
 EcoT22I  
 Ppu10I  
 BpmI  
 tgatctcacagccatctgtgatgcattcagaagcctgtgttaaatgcccttcttaggaaatgagctggagccacttgc  
 base pairs  
 actagagtgtcggtagacactacgtagtcttcggacacatttacgggaagatcctttactcgacctcgggtgaacg  
 3826 to 3900  
 NsiI  
 Zsp2I  
 GsuI  
 67/173  
 Asp700I  
 BsaMI  
 Mva1269I  
 ApoI  
 agaagataattctccaccaaaagccgaatatgaatgctgttatttctttacagaagatcattgaaatttcaaagtat  
 base pairs  
 tcttctataagagggtgttcgggcttatacttacgacaataaaagaatgtcttctagtaactttaagtttcata  
 3901 to 3975  
 XmnI  
 BsmI  
 AcsI

FIG. 12-27

MflI AccB1I  
 BstI BsaWI KpnI  
 BamHI BshNI  
 DraI  
 AvaI BcoI  
 MflI Eco88I PspAI  
 XhoII Cfr9I SmaI MslI  
 gtctttaagttctctggatccggtaccagattacaaggacgacgatgacaaagtagat cccgggtggcatccctg  
 base pairs  
 cagaaatttcaagagacctaggccatgggtctaatgttccctgctgctactgttcatcta gggccaccgtagggac  
 3976 to 4050  
 XhoII BanI Eco64I  
 BstYI Acc65I  
 BstX2I Asp718I  
 BstYI Ama87I  
 BstX2I BsoBI  
 XmaI PspAI  
 Eco130I  
 StyI  
 EcoT14I  
 GsuI  
 MslI  
 tgacccctccccagtgccctctcctggccttgggaagtggccactccagtggccaccagccttgtcctaataaaatt  
 base pairs  
 actggggaggggtcacggagagaccggaaccttcaacggtgagggtcacgggtgggtcggaacaggattattttaa  
 4051 to 4125  
 BssT1I  
 ErhI  
 BpmI

68/173

FIG. 12-28





70/173

BcoI  
Ama87I  
BcgI  
AvaI

tggtcactgcaatctcgcctcctgggttcaagcgattctcctgcctcagcctcccgagttgttgggattccag  
base pairs  
accgagtgcgttagagcgaggaccacaagttcgctaagaggacggagtcggagggctcaacaaccctaaggctc  
4276 to 4350

Eco88I  
BsoBI

NspI  
PaeI Mph1103I  
Ppu10I EcoT22I

BlpI

MscI  
MluNI  
EaeI

Esp3I

gcattgcattgacccaggtcagctcagctaatcttctgtttttgttagagacgggtttccaccataattggccaggctgggtc  
base pairs  
cgtagctactggtccgagtcgattaaaaaaccaaaaccatctctgccccaaagtgtataaccggctccgaccag  
4351 to 4425

BbuI Zsp2I  
SphI  
NsiI

CellII  
Bsp1720I  
Bpu1102I

BsmBI  
CfrI  
Bali

FIG. 12-30

Eco130I  
 StyI  
 EcoT14I  
 BstXI  
 BsaI  
 tccaaactcctaattcaggatgatctaccaccttggcctcccaaatgtggtgattacaggcgtgaaccactgct  
 base pairs  
 aggttgaggattagagtagatgggtggaaccggagggtttaacgaccctaattgtccgcacttggtgacga  
 4426 to 4500  
 Eco31I  
 BssT1I  
 ErhI

71/173

FIG. 12-31

BblII NcoI  
 HinII StyI  
 AclI AatII EcoT14I  
 DraI  
 cccttccctgtccttctgatttttaaaataactataccagcaggaggtccagacacagcataggctacctgcc  
 base pairs  
 gggaaggacaggaagactaaaattttattgatatggctcctcctgcaggctctgtgtcgatatccgatggacgg  
 4501 to 4575  
 Msp17I BssT1I  
 BsaHI ErhI  
 Hsp92I BspMI  
 72/173

Eco130I BsrFI PflMI  
 DsaI AgeI Bse118I  
 BsaWI AccB7I  
 atggcccaaccgggtgggacatttgagttgcttggtggcactgtcctctcatgcgttgggtccactcagtagatg  
 base pairs  
 taccgggttggccaccctgtaaaactcaacgaaccgtgacaggagagtagcgaacccaggtagtcatctac  
 4576 to 4650  
 BssAI Esp1396I  
 BstDSI PinAI Van91I  
 Bsp19I Cfr10I

FIG. 12-32

EaeI      AlwNI  
 cctgttgaattgggtacggccagcttctgtggaatgtgtcagttaggtgtggaaagtccccaggctcccc  
 base pairs  
 ggacaacttaacccatgcgcggtcggaagacaccttacacacagtcaatccacacacctttcaggggtccgagggg  
 4651 to 4725  
 CfrI  
 73/173  
 NspI  
 PaeI Mph1103I      SexAI  
 Ppu10I EcoT22I  
 agcaggcagaagtatgcaaaagcatgcattctcaattagtcagcaaccagggtgtggaaaagtccccagggtccccag  
 base pairs  
 tcgtccgtcttcatacgtttcgtacgtagagtaatacagtcggtgtccacacctttcaggggtccgaggggtc  
 4726 to 4800  
 BbuI Zsp2I  
 SphI  
 NsiI

FIG. 12-33

74/173

NspI  
PaeI Mph1103I  
Ppu10I EcoT22I  
caggcagaagtatgcaaaagcatgcatctcaattagtcagcaaccatagtcgccgccctaaactccgcccattccccgc  
base pairs  
gtccgtcttcatacgtttcgtacgtagagttaatcagtcgcttggtatcagggcggggattgagggcgggtaggggcg  
4801 to 4875

BbuI Zsp2I  
SphI  
NsiI

NcoI Bsp19I  
StyI BstDSI  
EcoT14I  
ccctaactccgcccagttccgcccattctccgcccattggtgactaattttttttattatgcagagggccgagg  
base pairs  
gggattgagggcgggtcaaggcgggtaagagggcgggtaccgactgattaaaaaaaaataaacgtctccgggtccc  
4876 to 4950

BssT1I  
ErhI Eco130I  
DsaI

FIG. 12-34

75/173

SseBI AvrII  
 Eco147I BlnI  
 StuI BstII  
 BseRI  
 BglI  
 ccgcctcgccctctgagctattccagaagtagtgaggaggttttttgaggcctaggcttttgcaaaaagctc c  
 base pairs  
 ggaggagccggagactcgataagggtcttcactcactcctccgaaaaaacctccggatccgaaaacgttttttcgagg  
 4951 to 5025  
 SfiI  
 AatI StyI  
 Pme55I ErhI  
 EcoT14I Eco130I  
 Ama87I  
 Eco88I BseRI  
 SfiI  
 ApoI  
 AvaI BsoBI  
 tcgaggaaactgaaaaaccagaaagttaattccctatagtgagtcgtattaaattcgtaatcatgggtcatagctgt  
 base pairs  
 agtccttgacttttttggtctttcaattaaggatatactcagcataaatttaagcattagtagcagtagtcgaca  
 5026 to 5100  
 XhoI BcoI  
 BstSFI  
 Acsi  
 Sfr274I  
 Paer7I

FIG. 12-35

76/173

AccBSI

BsrBI

ttcctgtgtgaaattgttatccgctcacaaattccacacaacatacagagcgggaagcataaaagtgtaaagcctggg  
base pairs  
aaggacacacttttaacaataggcgagtggttaaggtgtgtgtatgctcggccttcgtatttcacacatttcgggaccc  
5101 to 5175

BstD102I

AccB1I

BshNI

VspI

PshBI

gtgcctaatagtgagtgagctaactcacattaattgcgttgcgctcactgcccgcctttccagtcgggaaacctgtcgt  
base pairs  
caggattactcactcgattgagtgtaattaacgcaacgcgagtgacgggcgaaaggtcagccctttggacagca  
5176 to 5250

AsnI

AseI

BamI

Eco64I

FIG. 12-36

VspI  
 MspAII  
 PvuII PshBI  
 EaeI  
 gccagctgcattaatgaatcggccaacgcgcggggagagcggttgcgtattgggcgctcttcgcgttcctcgc  
 base pairs  
 cggtcgacgtaattacttagcgggttcgcgcggccctctccgccaacgcataaccgcgagaaggcggaaggagcg  
 5251 to 5325  
 NspBII  
 CfrI  
 HaeII EarI  
 Sapi  
 Ksp632I  
 77/173  
 BstH2I  
 Bsp143II  
 BstMCI  
 BsaOI  
 AccBSI  
 BsrBI  
 tcaactgactcgctcgctcggttcggctcggcgcgagcgggtatcagctcactcaaaggcggtaataacggttat  
 base pairs  
 agtgactgagcgacgcgagccagcaagccgacgcgcgtcgccatagtcgagtgagtttcgcgcattatgcgaata  
 5326 to 5400  
 Bsh1285I  
 BsiEI  
 BstD102I

FIG. 12-37



78/173

NspI

BspLU11I

ccacagaatcagggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaagg  
base pairs  
gggtcttagtcccctattgcgtcctttctgtacactcgttttcggtcgttttcggtccttggcattttttcc  
5401 to 5475

AflIII

DrdI

ccgcgttgctggcggttttccataggctccgccccctgacgagcatcacaaaaatcgacggtcaagtcagaggt  
base pairs  
ggcgcaacgacccgcaaaaaggatccgaggggggactgctcgtagtggttttagctgcgaggttcagtcctcca  
5476 to 5550

FIG. 12-38

BsiI

ggcgaaacccgacaggactataaagataccaggcggtttccccctggaagctccctcgtgcgctctcctgttccga  
 base pairs  
 ccgctttgggctgtcctgatatattctatggtccgcaaaaggggaccttcgagggagcacgcgagaggacaaggct  
 5551 to 5625

BssSI

79/173

BsaWI                      BstH2I                      SfcI

ccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgta  
 base pairs  
 gggacggcgcaatggcctatggacaggcggaagagggaagcccttcgcaccgcgaaagagttacgagtgcgacat  
 5626 to 5700

HaeII                      BstSFI

FIG. 12-39

BsiHKAI	NspBII
Alw44I	BstMCI
VneI Bbv12I	BsaOI
ggatatcagttcgggtgtaggtcggttcggtccaagctgggctgtgtgcacgaaccccccggttcagccccgaccgct	
base pairs	
ccatagagtcaggccacatccagcaagcgaggttcgacccgacacacacgtgcttggggggcaagtcgggctggcga	
5701 to 5775	
ApaLI	Bsh1285I
AspHI	BsiEI
Alw21I	MspAII
	80/173

BsaWI	AlwNI
gagccttatccggtaactatcgctcttgagtcgaacccggtaagacacgacttatcgccactggcagcagccactg	
base pairs	
cgcggaataggccattgatagcagaactcaggttggccattctgtgctgaatagcggtgaccgtcgctcggtgac	
5776 to 5850	

FIG. 12-40

81/173

## SfCI

gtaacaggattagcagagcgaggatgtaggcgggtgctacagagttcttgaagtggcctaactacggctaca  
base pairs  
cattgtcctaatacgtctcgctccatacatccgccacgatgtctcaagaacttcaccaccggattgatgccgatgt  
5851 to 5925

## BstSFI

## Eco57I

ctagaagaacagtatatttggtatctgcgctctgctgaagccaggttaccttcggaaaaagagttggtagctcttgat  
base pairs  
gatcttcttgtcataaaccatagacgcgagacgacttcggtcaatggaagcctttttctcaaccatcgagaacta  
5926 to 6000

FIG. 12-41

MflI  
 XhoII  
 NspBII  
 ccggcaaaaccaccgctggtagcgggtgtttttgttgcaagcagcagattacgcgcagaaaaaaggat  
 base pairs  
 ggccgtttgtttggtggcgaccatcgccaccaaaaaaacggttcgtcctaatagcgcgtcttttttccta  
 6001 to 6075  
 MspAII  
 BstYI  
 BstX2I

82/173

MflI  
 XhoII  
 ctcaagaagatcctttgatctttctacggggtctgacgctcagtggaacgaaaaactcacggttaaggatttgg  
 base pairs  
 gagttcttctaggaaactagaaaaagatgccccagactgcgagtcaccttgcttttgagtgcattccctaaaaacc  
 6076 to 6150  
 BstYI  
 BstX2I

FIG. 12-42

83/173

MflI MflI MflI DraI  
 XhoII XhoII XhoII DraI  
 RcaI  
 tcatgagattatcaaaaaggatccttcacctagatccttttaaaattaaaaatgaagtttttaaatcaatctaaagta  
 base pairs  
 agtactctaataagtttttccctagaagtgatctaggaataatttaatttttacttcaaaatttagtttagatttcat  
 6151 to 6225  
 BspHI BstYI BstYI  
 BstX2I BstX2I

AccB1I  
 BshNI  
 tatatgagtaaacttggtctgacagttaccaaatgcttaatcagtgaggcacctatctcagcgatctgtctatttc  
 base pairs  
 atatactcatttgaaccagactgtcaatgggttacgaattagtcactccgtggatagatcgctagacagataaag  
 6226 to 6300  
 BanI  
 Eco64I

FIG. 12-43

Eam1105I  
 AspEI  
 gttcatccatagttgcctgactccccgtcggtagataactacgatacggagggttaccatctggccccagtg  
 base pairs  
 caagtaggtatcaacggactgaggggagcacatctattgatgctatgccctcccgaatggtagaccgggggtcac  
 6301 to 6375  
 EclHKI  
 AhdI  
 84/173  
 Cfr10I  
 BsrDI BsaI BssAI BpmI BglI  
 ctgcaatgataccgcgagagaccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccg  
 base pairs  
 gacgttactatggcgctctgggtgcgagtggccgaggtctaaatagtcgttatttggtcggcggccttccccggc  
 6376 to 6450  
 Eco31I BsrFI GsuI  
 Bse118I

FIG. 12-44

VspI  
PshBI

agcgagaaagtggtcctgcaactttatccgcctccatccagtcatttaattggtgccgggaagctagagtaagta  
base pairs  
tcgcgtcttcaccaggacggttgaaataggcggaggttaggtcagataattaacaacggcccttcgatctcattcat  
6451 to 6525

AsnI  
AseI

85/173

AviII  
FspI

BstSFI  
SfcI

MslI

gttcgccagttaatagtttgcgcaacggttggtgccattgctacaggcatcggtgtcacgctcgctgttggtta  
base pairs  
caagcgggtcaattatcaaacggttgcaacaacggtaacgatgtccgtagcaccacagtgcgagcagcaaacat  
6526 to 6600

Acc16I  
Psp1406I

BsrDI

FIG. 12-45



BsaWI  
 tggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcgggtta  
 base pairs  
 accgaagtaagtcgagggccaagggttgctagtccgctcaatgtactaggggtacaacacgttttttcgccaat  
 6601 to 6675

86/173

BstMCI	EaeI	MslI
PvuI BsiEI		
BsaOI		
gctccttcggtcctccgatcgttggtcagaagtaagttggccgcagtggttatcactcatggttatggcagcactgc		
base pairs		
cgaggaagccaggaggtagcaacagtccttcattcaaccggcggtcacaatagtgagtacccaataccgtcgtgacg		
6676 to 6750		
BspCI	CfrI	
Bsh1285I		
Ple19I		

FIG. 12-46

Acc113I  
 Eco255I  
 ataattcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtactcaaccaagtcattcttgag  
 base pairs  
 tattaagagaatgacagtagcggtaggcattctacgaaaagacactgaccactcatgagttgggttcagtaagactc  
 6751 to 6825  
 ScaI

87/173

BstMCI      BbiII  
 BsaOI      BcgI      HinII  
 AclI  
 aatagtgtatgcggcgaccgagttgctcttgcccgcggtcaatacgggataataccgcgccacatagcagaactt  
 base pairs  
 ttatcacatacgcgctggctcaacgagaaacgggcccaggttatgccctattatggcgcggtgtatcgctcttgaa  
 6826 to 6900  
 Bsh1285I      Msp17I  
 BsiEI      BsaHI  
 Hsp92I

FIG. 12-47

88/173

Alw21I	XmnI	MflI	MflI
DraI	Psp1406I	XhoII	NspBII XhoII
taaaagtgcctcatcattgggaaaaacgttcttcggggcgaaaaactctcaaggatcttaccgctgttgagatccagtt			
base pairs			
atttcacgagtagtaaccttttgcagaagccccgcttttgagagttcctagaatggcgacaactctaggtcaa			
6901 to 6975			
BsiHKAI	Asp700I	BstYI	MspAII BstYI
Bbv12I		BstX2I	BstX2I

BssSI	
Alw44I Bbv12I	
VneI BsiHKAI	Eco57I
cgatgtaaccactcgtgcaccaccaactgatcttcagcatcttttactttcaccagcgtttctgtgggtgagcaaaaa	
base pairs	
gctacattgggtgagcacgtgggttgactagaagtcgtagaaaaatgaaagtggtcgcaaaagaccactcggtttt	
6976 to 7050	
ApalI Alw21I	
BsiI	
AspHI	

FIG. 12-48

EarI  
 MslI  
 Fam1104I  
 caggaaggcaaaatgccgcaaaaaagggaataaggcgacacgggaatgttgaatactcatactcttccttttc  
 base pairs  
 gtccttcgcgtttttacggcggtttttcccttattcccgctgtgcctttacaacttatgagtatgagaaggaaaaag  
 7051 to 7125  
 Ksp632I

89/173

SspI  
 RcaI  
 AccBSI  
 BsrBI  
 aatattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaac  
 base pairs  
 ttataataacttcgtaaatagtcaccaataacagagtagtgcctatgtataaaacttacataaatctttttatttg  
 7126 to 7200  
 BspHI  
 BstD102I

FIG. 12-49

SfcI

aaataggggttccgcgcacatttccccgaaaagtgccaccctgacgcgccctgtagggcgcatatagcgcgcgccg  
 base pairs  
 ttatccccaaaggcgcgtgtaaaagggctttcacgggtggactgcgcgggacatcgccgctaattcgcgcgcgc  
 7201 to 7275

BstSFI

90/173

AccBSI

BstH2I   HaeII   BstD102I

Bsp143II   BsrBI

gtgtggtgttacgcgcagcgtgaccgctacacttgccagcgccttagcgcgcccttctcgcttcttccctt  
 base pairs  
 cacaccaccaatgcgcgtcgcaactggcgatgtgaacggtcgcgggtcgcgggcgaggaagcgaaaaggggaa  
 7276 to 7350

HaeII   Bsp143II

BstH2I

FIG. 12-50

91/173

BsrFI  
BssAI NaeI  
MroNI Bse118I  
cctttctgccacgttcgccggctttccccgtcaagctctaaatcggggcatccctttagggttccgatttagtg  
base pairs  
ggaaagagcgggtgcaagcgccgaaaggcagtttcgagatttagccccgtaggggaaatcccaaggctaaatcac  
7351 to 7425  
NgoAIV  
NgomI  
Cfr10I

AccB1I  
BshNI  
BsaAI  
ctttacggcacctcgacccccaaaaacttgattagggatggttcacgtagtgggccatcgccctgatagacgg  
base pairs  
gaaatgccgtggagctgggggttttttgaactaatcccactaccaagtgcaccocggtagcgggactatctgcc  
7426 to 7500  
BamI  
Eco64I  
DraIII

FIG. 12-51

92/173

## DrdI

tttttcgccctttgacgttggagtcacggttccttaatagtggactcttggtccaaactggaacaactcaacc  
base pairs  
aaaaagcgggaaactgcaacctcaggtgcaagaaattatcacctgagaacaaggtttgaccttggttgagttgg  
7501 to 7575

ctatctcgggtctattcttttgatttataagggttttgccgatttcggcctattgggttaaaaatgagctgattt  
base pairs  
gataagagccagataagaaaaactaaatatccctaaaacgggtaaagccggataaaccaattttttactcgactaaa  
7576 to 7650

FIG. 12-52

ApoI                      SspI    Psp1406I  
 aacaaaaatttaacgcgaattttaacaaaaataattaaacgtttacaattt    base pairs  
 ttgtttttaaatgctgaatttaaaattgttttataaatttgcaaatgttaaa    7651 to 7699  
 ACSI                      ACSI

Table by Enzyme Name

Enzyme name	No. cuts	Positions of sites	Recognition sequence
AatI	3	3446 3546 5002	agg/cct <u>More info</u>
AatII	5	451 504 587 773 4550	gacgt/c <u>More info</u>
Acc113I	1	6804	agt/act <u>More info</u>
Acc16I	2	21 6546	tgc/gca <u>More info</u>
Acc65I	3	2264 3434 3998	g/ gtacc <u>More info</u>
AccB1I	8	791 2264 3065 3434 3998 5175 6272 7432	g/ gyrcc <u>More info</u>
AccB7I	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg <u>More info</u>
AccBSI	4	5126 5367 7168 7332	gagcgg <u>More info</u>
AclNI	1	326	a/ ctagt <u>More info</u>
AcSI	8	912 1990 2244 2994 3963 5075 7656 7667	r/ aatty <u>More info</u>
AcyI	6	448 501 584 770 4547 6861	gr/cgyc <u>More info</u>

FIG. 12-53

93/173



AflIII	3	2702 3796 5431	a/ crygt	<u>More info</u>
AgeI	1	4584	a/ ccggt	<u>More info</u>
AhdI	2	4150 6324	gacnnn/nngtc	<u>More info</u>
Alw21I	6	894 1576 2330 5749 6910 6995	gwgwc/c	<u>More info</u>
Alw44I	2	5745 6991	g/ tgcac	<u>More info</u>
AlwNI	6	1147 1273 1775 3091 4678 5847	cagnnn/ctg	<u>More info</u>
Ama87I	3	4034 4330 5025	c/ ycgrg	<u>More info</u>
AocI	3	1034 1046 3256	cc/ tnagg	<u>More info</u>
Apal	1	4202	gggcc/c	<u>More info</u>
ApalI	2	5745 6991	g/ tgcac	<u>More info</u>
ApoI	8	912 1990 2244 2994 3963 5075	r/ aatty	<u>More info</u>
		7656 7667		
AseI	4	334 5202 5261 6496	at/ taat	<u>More info</u>
AsnI	4	334 5202 5261 6496	at/ taat	<u>More info</u>
Asp700I	5	1107 2481 3506 3906 6923	gaann/nnttc	<u>More info</u>
Asp718I	3	2264 3434 3998	g/ gtacc	<u>More info</u>

FIG. 12-54

94/173

AspEI	2	4150	6324					gacnnn/nngtc	<u>More info</u>
AspHI	6	894	1576	2330	5749	6910	6995	gwgcw/c	<u>More info</u>
AspI	1	3674						gacn/nngtc	<u>More info</u>
AtsI	1	3674						gacn/nngtc	<u>More info</u>
AvaI	3	4034	4330	5025				c/ ycgrg	<u>More info</u>
AviII	2	21	6546					tgc/gca	<u>More info</u>
AvrII	2	3109	5003					c/ ctagg	<u>More info</u>
BalI	5	184	238	3300	3653	4414		tgg/cca	<u>More info</u>
BamHI	1	3992						g/ gatcc	<u>More info</u>
BanI	8	791	2264	3065	3434	3998	5175	g/ gyrcc	<u>More info</u>
		6272	7432						95/173
BanII	5	894	1017	1623	3526	4202		grgcy/c	<u>More info</u>
BanIII	1	939						at/ cgat	<u>More info</u>
BbiII	6	448	501	584	770	4547	6861	gr/cgyc	<u>More info</u>
BbrPI	1	2705						cac/gtg	<u>More info</u>
BbsI	2	2512	4216					gaagac	<u>More info</u>
BbuI	4	2930	4355	4750	4823			gcatg/c	<u>More info</u>
Bbv12I	6	894	1576	2330	5749	6910	6995	gwgcw/c	<u>More info</u>
Bbv16II	2	2512	4216					gaagac	<u>More info</u>
BcgI	4	941	2556	4321	6851			cgannnnntgc	<u>More info</u>

FIG. 12-55

BclI	1	969		t/ gatca	<u>More info</u>
BcoI	3	4034 4330 5025		c/ ycgrg	<u>More info</u>
BglI	5	14 417 538 4956 6444		gccnnnn/nggc	<u>More info</u>
BglII	2	932 3409		a/ gatct	<u>More info</u>
BlnI	2	3109 5003		c/ ctagg	<u>More info</u>
BlpI	3	1200 2337 4366		gc/tnagc	<u>More info</u>
BpiI	2	2512 4216		gaagac	<u>More info</u>
BpmI	10	1015 1279 1772 2781 2842 3022		ctggag	<u>More info</u>
		3892 4097 4259 6414			
Bpu1102I	3	1200 2337 4366		gc/tnagc	<u>More info</u>
Bpu14I	3	1603 1988 2423		tt/cgaa	<u>More info</u>
BpuAI	2	2512 4216		gaagac	<u>More info</u>
Bsa29I	1	939		at/ cgat	<u>More info</u>
BsaAI	3	666 2705 7473		yac/gtr	<u>More info</u>
BsaHI	6	448 501 584 770 4547 6861		gr/cgyc	<u>More info</u>
BsaI	3	3380 4427 6396		ggtctc	<u>More info</u>
BsaMI	3	1886 3631 3936		gaatgc	<u>More info</u>
BsaOI	7	42 424 928 5347 5771 6694 6843		cgry/cg	<u>More info</u>
BsaWI	6	3200 3995 4584 5637 5784 6615		w/ ccggw	<u>More info</u>
BsCI	1	939		at/ cgat	<u>More info</u>

96/173

FIG. 12-56

97/173

Bse118I	3	4584	6404	7368	r/ ccggy	<u>More info</u>
Bse21I	3	1034	1046	3256	cc/ tnagg	<u>More info</u>
BseCI	1	939			at/ cgat	<u>More info</u>
BseRI	5	1337	1671	3725 4989 5027	gaggag	<u>More info</u>
BsgI	3	2315	3212	4264	gtgcag	<u>More info</u>
Bsh1285I	7	42	424	928 5347 5771 6694 6843	cgry/cg	<u>More info</u>
BshNI	8	791	2264	3065 3434 3998 5175	g/ gyrcc	<u>More info</u>
		6272	7432			
BsiEI	7	42	424	928 5347 5771 6694 6843	cgry/cg	<u>More info</u>
BsiHKAI	6	894	1576	2330 5749 6910 6995	gwgw/c	<u>More info</u>
BsiI	2	5609	6993		ctcgtg	<u>More info</u>
BsmBI	3	2023	2773	4397	cgtctc	<u>More info</u>
BsmI	3	1886	3631	3936	gaatgc	<u>More info</u>
BsoBI	3	4034	4330	5025	c/ ydgrg	<u>More info</u>
Bsp106I	1	939			at/ cgat	<u>More info</u>
Bsp119I	3	1603	1988	2423	tt/cgaa	<u>More info</u>
Bsp120I	1	4198			g/ ggccc	<u>More info</u>
Bsp1407I	2	270	3471		t/ gtaca	<u>More info</u>
Bsp143II	5	2519	5309	5679 7318 7326	rgcgc/y	<u>More info</u>
Bsp1720I	3	1200	2337	4366	gc/tnagc	<u>More info</u>
Bsp19I	6	686	3324	3424 3600 4574 4910	c/ catgg	<u>More info</u>

FIG. 12-57

98/173

BspCI	2	42 6694	cgat/cg	<u>More info</u>
BspDI	1	939	at/ cgat	<u>More info</u>
BspHI	3	1891 6151 7159	t/ catga	<u>More info</u>
BspLU11I	1	5431	a/ catgt	<u>More info</u>
BspMI	2	1913 4574	acctgc	<u>More info</u>
BspXI	1	939	at/ cgat	<u>More info</u>
BsrBI	4	5126 5367 7168 7332	gagcgg	<u>More info</u>
BsrDI	4	245 2827 6383 6565	gcaatg	<u>More info</u>
BsrFI	3	4584 6404 7368	r/ ccggy	<u>More info</u>
BsrGI	2	270 3471	t/ gtaca	<u>More info</u>
BssAI	3	4584 6404 7368	r/ ccggy	<u>More info</u>
BssSI	2	5609 6993	ctcgtg	<u>More info</u>
BssT1I	13	686 1950 2226 3109 3324 3424 3547 3600 4077 4456 4574 4910 5003	c/ cwwgg	<u>More info</u>
BstBI	3	1603 1988 2423	tt/cgaa	<u>More info</u>
BstD102I	4	5126 5367 7168 7332	gagcgg	<u>More info</u>
BstDSI	7	686 1062 3324 3424 3600 4574 4910	c/ crygg	<u>More info</u>
BstH2I	5	2519 5309 5679 7318 7326	rgcgc/y	<u>More info</u>

FIG. 12-58

BstI	1	3992	g/ gatcc	<u>More info</u>
BstMCI	7	42 424 928 5347 5771 6694 6843	cgry/cg	<u>More info</u>
BstSFI	8	944 2144 4220 5058 5696 5887	c/ tryag	<u>More info</u>
		6565 7250		
BstSNI	1	666	tac/gta	<u>More info</u>
BstX2I	12	932 2400 2634 3409 3992 4030	r/ gatcy	<u>More info</u>
		6072 6083 6169 6181 6949 6966		
BstXI	3	3076 3325 4473	ccannnnn/ntgg	<u>More info</u>
BstYI	12	932 2400 2634 3409 3992 4030	r/ gatcy	<u>More info</u>
		6072 6083 6169 6181 6949 6966		
BstZI	1	925	c/ ggccg	<u>More info</u>
Bsu15I	1	939	at/ cgat	<u>More info</u>
Bsu36I	3	1034 1046 3256	cc/ tnagg	<u>More info</u>
CciNI	1	925	gc/ggccgc	<u>More info</u>
CelII	3	1200 2337 4366	gc/tnagc	<u>More info</u>
Cfr10I	3	4584 6404 7368	r/ ccggy	<u>More info</u>
Cfr9I	1	4034	c/ ccggg	<u>More info</u>
CfrI	10	152 182 236 925 3298 3651 4412	y/ ggccr	<u>More info</u>
		4669 5270 6712		
Clal	1	939	at/ cgat	<u>More info</u>
Csp45I	3	1603 1988 2423	tt/cgaa	<u>More info</u>
CvnI	3	1034 1046 3256	cc/ tnagg	<u>More info</u>

99/173

FIG. 12-59

DraI	5	3981	4523	6190	6209	6901	ttt/aaa	<u>More info</u>
DraII	3	3291	4198	4225			rg/gnccy	<u>More info</u>
DraIII	1	7476					cacnnn/gtg	<u>More info</u>
DrdI	3	1076	5539	7520			gacnnnn/nngtc	<u>More info</u>
DsaI	7	686	1062	3324	3424	3600 4574	c/ crygg	<u>More info</u>
		4910						
EaeI	10	152	182	236	925	3298 3651 4412	y/ ggccr	<u>More info</u>
		4669	5270	6712				<u>More info</u>
EagI	1	925					c/ ggccg	<u>More info</u>
Eam1104I	5	58	2482	2793	5314	7118	ctcttc	<u>More info</u>
Eam1105I	2	4150	6324				gacnnn/nngtc	<u>More info</u>
EarI	5	58	2482	2793	5314	7118	ctcttc	<u>More info</u>
Ecl1136II	1	892					gag/ ctc	<u>More info</u>
EclHKI	2	4150	6324				gacnnn/nngtc	<u>More info</u>
EclXI	1	925					c/ ggccg	<u>More info</u>
Eco105I	1	666					tac/gta	<u>More info</u>
Eco130I	13	686	1950	2226	3109	3324 3424	c/ cwwgg	<u>More info</u>
		3547	3600	4077	4456	4574 4910		
		5003						
Eco147I	3	3446	3546	5002			agg/cct	<u>More info</u>

100/173

FIG. 12-60

101/173

Eco24I	5	894	1017	1623	3526	4202	grgcy/c	<u>More info</u>
Eco255I	1	6804					agt/act	<u>More info</u>
Eco31I	3	3380	4427	6396			ggtctc	<u>More info</u>
Eco32I	1	952					gat/ atc	<u>More info</u>
Eco52I	1	925					c/ ggccg	<u>More info</u>
Eco57I	7	1210	2446	2488	3271	3314 5963	ctgaag	<u>More info</u>
		7011						
Eco64I	8	791	2264	3065	3434	3998 5175	g/ gyrcc	<u>More info</u>
		6272	7432					
Eco72I	1	2705					cac/gtg	<u>More info</u>
Eco81I	3	1034	1046	3256			cc/ tnagg	<u>More info</u>
Eco88I	3	4034	4330	5025			c/ ycgrg	<u>More info</u>
EcoICRI	1	892					gag/ ctc	<u>More info</u>
EcoNI	4	1259	1338	1684	3723		cctnn/nnnagg	<u>More info</u>
EcoO109I	3	3291	4198	4225			rg/gnccy	<u>More info</u>
EcoRI	3	912	1990	2994			g/ aattc	<u>More info</u>
EcoRV	1	952					gat/ atc	<u>More info</u>
EcoT14I	13	686	1950	2226	3109	3324 3424	c/ cwwgg	<u>More info</u>
		3547	3600	4077	4456	4574 4910		
		5003						
EcoT22I	5	3703	3850	4357	4752	4825	atgca/t	<u>More info</u>

FIG. 12-61



ErhI	13	686 1950 2226 3109 3324 3424	c/ cwwgg	<u>More info</u>
		3547 3600 4077 4456 4574 4910		
		5003		
Esp1396I	6	1445 1482 1775 1796 2644 4587	ccannnn/ntgg	<u>More info</u>
Esp3I	3	2023 2773 4397	cgtctc	<u>More info</u>
FauNDI	1	560	ca/ tatg	<u>More info</u>
FbaI	1	969	t/ gatca	<u>More info</u>
FriOI	5	894 1017 1623 3526 4202	grgcy/c	<u>More info</u>
Fspi	2	21 6546	tgc/gca	<u>More info</u>
GsuI	10	1015 1279 1772 2781 2842 3022	ctggag	<u>More info</u>
		3892 4097 4259 6414		
HaeII	5	2519 5309 5679 7318 7326	rgcgc/y	<u>More info</u>
HinII	6	448 501 584 770 4547 6861	gr/cgyc	<u>More info</u>
HincII	3	311 446 842	gty/rac	<u>More info</u>
HindII	3	311 446 842	gty/rac	<u>More info</u>
HindIII	3	918 1394 2183	a/ agctt	<u>More info</u>
Hsp92I	6	448 501 584 770 4547 6861	gr/cgyc	<u>More info</u>
KpnI	3	2268 3438 4002	ggtac/c	<u>More info</u>
Ksp22I	1	969	t/ gatca	<u>More info</u>
Ksp632I	5	58 2482 2793 5314 7118	ctcttc	<u>More info</u>
LspI	3	1603 1988 2423	tt/cgaa	<u>More info</u>
MfeI	2	1091 3773	c/ aattg	<u>More info</u>
MflI	12	932 2400 2634 3409 3992 4030	r/gatcy	<u>More info</u>

102/173

FIG. 12-62

103/173

MluNI	5	6072 6083 6169 6181 6949 6966	tgg/cca	<u>More info</u>
Mph1103I	5	184 238 3300 3653 4414	atgca/t	<u>More info</u>
MronI	1	3703 3850 4357 4752 4825	g/ ccggc	<u>More info</u>
MSCI	5	184 238 3300 3653 4414	tgg/cca	<u>More info</u>
MslI	10	691 2094 2703 3323 3489 4047	caynn/nnrtg	<u>More info</u>
		4094 6576 6735 7094		
Msp17I	6	448 501 584 770 4547 6861	gr/cgyc	<u>More info</u>
MspAII	7	71 2341 2731 5255 5773 6018 6959	cmg/ckg	<u>More info</u>
MunI	2	1091 3773	c/ aattg	<u>More info</u>
Mva1269I	3	1886 3631 3936	gaatgc	<u>More info</u>
NaeI	1	7370	gcc/ggc	<u>More info</u>
NcoI	6	686 3324 3424 3600 4574 4910	c/ catgg	<u>More info</u>
NdeI	1	560	ca/ tatg	<u>More info</u>
NgoAIV	1	7368	g/ ccggc	<u>More info</u>
NgomI	1	7368	g/ ccggc	<u>More info</u>
NotI	1	925	gc/ggccgc	<u>More info</u>
NsiI	5	3703 3850 4357 4752 4825	atgca/t	<u>More info</u>
NspBII	7	71 2341 2731 5255 5773 6018 6959	cmg/ckg	<u>More info</u>
Nspi	5	2930 4355 4750 4823 5435	rcatg/y	<u>More info</u>
NspV	3	1603 1988 2423	tt/cgaa	<u>More info</u>
PaeI	4	2930 4355 4750 4823	gcatg/c	<u>More info</u>
Paer7I	1	5025	c/ tcgag	<u>More info</u>

FIG. 12-63

104/173

PfIMI	6	1445	1482	1775	1796	2644	4587	ccannnn/ntgg	More info
PinAI	1	4584						a/ ccggt	More info
Ple19I	2	42	6694					cgat/cg	More info
PmaCI	1	2705						cac/gtg	More info
Pme55I	3	3446	3546	5002				agg/cct	More info
PmlI	1	2705						cac/gtg	More info
Ppu10I	5	3699	3846	4353	4748	4821		a/ tgcac	More info
PshBI	4	334	5202	5261	6496			at/ taat	More info
Psp124BI	1	894						gagct/c	More info
Psp1406I	3	6550	6923	7687				aa/cggt	More info
PspAI	1	4034						c/ccggg	More info
PspALI	1	4036						ccc/ggg	More info
PspOMI	1	4198						g/ggccc	More info
PstI	2	948	2148					ctgca/g	More info
PvuI	2	42	6694					cgat/cg	More info
PvuII	3	71	2341	5255				cag/ctg	More info
RcaI	3	1891	6151	7159				t/catga	More info
SacI	1	894						gagct/c	More info
SapI	2	2483	5314					gctcttc	More info
ScaI	1	6804						agt/act	More info
SexAI	1	4769						a/ ccwggg	More info
SfCI	8	944	2144	4220	5058	5696	5887	c/ tryag	More info

FIG. 12-64

SfiI	1	6565	7250	ggcnnnn/nggcc	<u>More info</u>
Sfr274I	1	4956		c/ tgcag	<u>More info</u>
SfuI	3	5025		tt/cgaa	<u>More info</u>
SmaI	1	1603	1988 2423	ccc/ggg	<u>More info</u>
SnaBI	1	4036		tac/gta	<u>More info</u>
SpeI	1	666		a/ ctagt	<u>More info</u>
SphI	4	326		gcatg/c	<u>More info</u>
SseBI	3	2930	4355 4750 4823	agg/cct	<u>More info</u>
SspBI	2	3446	3546 5002	t/ gtaca	<u>More info</u>
SspI	6	270	3471	aat/att	<u>More info</u>
SstI	1	179	226 3571 4164 7128 7681	gagct/c	<u>More info</u>
StuI	3	894		agg/cct	<u>More info</u>
StyI	13	3446	3546 5002	c/ cwwgg	<u>More info</u>
		686	1950 2226 3109 3324 3424		
		3547	3600 4077 4456 4574 4910		
		5003			
Tth111I	1	3674		gacn/nngtc	<u>More info</u>
Van91I	6	1445	1482 1775 1796 2644 4587	ccannnn/ntgg	<u>More info</u>
VneI	2	5745	6991	g/ tgcac	<u>More info</u>
VspI	4	334	5202 5261 6496	at/ taat	<u>More info</u>
XbaI	1	3811		t/ ctaga	<u>More info</u>
XcmI	2	1948	2897	ccannnn/nnnttgg	<u>More info</u>

105/173

FIG. 12-65

XhoI	1	5025				c/ tcgag	<u>More info</u>
XhoII	12	932	2400	2634	3409	3992	<u>More info</u>
			6072	6083	6169	6181	6949
XmaI	1	4034				c/ ccggg	<u>More info</u>
XmaIII	1	925				c/ ggccg	<u>More info</u>
XmnI	5	1107	2481	3506	3906	6923	<u>More info</u>
Zsp2I	5	3703	3850	4357	4752	4825	<u>More info</u>

106/173

The following endonucleases were selected but don't cut this sequence:

AccI, AccIII, AfeI, AflII, Aor51HI, AscI, BbeI, BfrI, BsaBI, Bse8I, BseAI, BsePI, Bsh1365I, BsiMI, BsiWI, Bsp13I, Bsp68I, BspEI, BspTI, BsrBRI, BssHII, Bst1107I, Bst98I, BstEII, BstPI, Cfr42I, CpoI, CspI, Eco47III, Eco91I, EcoO65I, EheI, FseI, HpaI, Kasi, Kpn2I, KspI, Mami, MluI, MroI, MspCI, NarI, NheI, NruI, PacI, Pfl23II, PmeI, PpuMI, PshAI, Psp5II, PspEI, PspLI, PstNHI, RsrII, SacII, Sali, SbfI, Sfr303I, Sgfi, SgrAI, SmiI, Sphi, Srfl, Sse8387I, SstII, SunI, SwaI, Vha464I

FIG. 12-66

FIG. 13A
FIG. 13B
FIG. 13C
FIG. 13D
FIG. 13E

ccattgccattcaggctgcgcaactgttgggaaggcgatcggcgggcctcttcgctattaccgagctggcgaaagggg  
ggatgtgctgcaaggcgattaagtggtaacgccagggtttccagtcacgacggttgtaaaacgacggccagtgccaagct  
gatctaataatattggccattagccatattattcattgggtatatagcataaataatattggctattggccattgcatacgttgatcca  
tatcataatatgtacatttatattggctcatgtccaacattaccgccatgttgacattgattattgactagttattaatagtaataatcaattacg  
gggtcattagttcatagccatatatggagttccgcgttacataacttacggtaaatggcccgtggcgaccccgagcgaccc  
ccgcccgttgacgtcaatagtgacgtatgttcccatagtaacgccaataggggactttccattgacgtcaatgggtggagttattacg  
gtaaaactgcccacttggcagtagcatcaagtgtagtataatgccaaagtcgcccctattgacgtcaatgacggtaaatggcccgcct  
agcattatgcccagtagacaccttacgggagtttccctacttggcagtagcatctacgtattagtcacgctattaccatgggtgatgcg  
gtttggcagtacaccaatggcggtggatagcgggttgactcacggggatttccaagtcctccacccattgacgtcaatgggaggtt  
tgthttggcaccaaaataacgggactttccaaaatgtcgtataacccccggccgttgacgcaaatggggcggtagggcggtgtacg  
gtgggagggtctatataagcagagctcgttttagtgaaaccgtcagaattcaagcttgcggccgcagatctatcgatctgcaggatatac  
(EcoRV)  
acc

107/173

FIG. 13A

FIG. 13

108/173

ATGCACAGTATGATCAGCTCAGTGGATGTGAAGTCAGAAATTCCCTGTGGGCCCTGGAGCCCATCTCACCTTTA  
GACCTAAGGACAGACCTCAGGATGATGATGCCCGTGGACCCCTGTTGTCCGTGAGAAAGCAATTGCAGCAG  
GAATTACTTCTTATCCAGCAGCAGCAACAATCCAGAAAGCAGCTTCTGATAGCAGAGTTTCAGAAACAGCAT  
GAGAACTTGACACGGCAGCACCAAGGCTCAGCTTCAGGAGCATATCAAGGAACTTCTAGCCATAAAACAGCAA  
CAAGAACTCCTAGAAAAGGAGCAGAAACTGGAGCAGCAGAGGCAAGAACAGGAAGTAGAGAGGCATCGCAGA  
GAACAGCAGCTTCCCTCTCAGAGGCAAGATAGAGGACGAGAAAGGCGAGTGGCAAGTACAGAAAGTAAAG  
CAGAAAGCTTCAAGAGTTCCCTACTGAGTAAATCAGCAACGAAAGACACTCCAACTAATGGAAAAAATCATTTCC  
GTGAGCCGCCATCCCAAGCTCTGGTACACGGCTGCCCAACCACACATCATTTGGATCAAAGCTCTCCACCCCTT  
AGTGGAAACATCTCCATCCTACAAGTACACATTAACAGGAGCAAGATGCAAGGATGATTTCCCCCTTCGA  
AAAACTGCCCTCTGAGCCCAAATTGAAGGTGGTCCAGGTTAAAAACAGAAAGTGGCAGAGAGGAGAAAGCAGC  
CCCTTACTCAGGCGGAAGGATGGAATGTGTCACTTCATTCAAGAAAGCGAATGTTTGAGGTGACAGAAATCC  
TCAGTCAGTAGCAGTTCTCCAGGCTCTGGTCCAGTTCACCAAACTGGCCCACTGGAAAGTGTACTGAA  
AATGAGACTTCGGTTTTGCCCCCTACCCCTCATGCGCAGCAAAATGGTTTCACAGCAACGCAATCTAAATTCA  
GAGATTCCATGAACCTGCTAAGTCTTTATACCTCTCCTTCTTTGCCCAACATFACCTTGGGGCTTCCCGCA  
GTGCCATCCCAGCTCAATGCTTTCGAATTCACTCAAAAGAAAGCAGAAAGTGTGAGACGCGCAGCTTAGGCAA  
GGTGTTCCTCTGCCCTGGCAGTATGGAGGCAGCATCCCGGCATCTTCCAGCCACCCCTCATGTTAATTAGAG  
GAAAGCCACCCAAACAGCAGCCACAGGCTCTCCTGCAGCATTTATTTATTTGAAAGAACAAATGCGACAGCAA  
AAGCTTCTTGTAGCTGGTGGAGTTCCCTTACATCCTCAGTCTCCCTTGGCAACAAAGAGAGAAATTTCACCT  
GGCATTAGAGGTACCCACAAATTGCCCCCGTCAAGACCCCTGAACCCAGTCTGCACCTTTTGCCCTCAG  
AGCACGTTGGCTCAGCTGGTCAATTCAAACAGCAACACAGCAATCTTGAGAAAGCAGAAAGCAATACCAGCAG  
CAGATCCACATGAACAAACTGCTTTCGAAAATCTATTGAACAACTGAAGCAACCCAGGAGTCACTTGTAGGAA  
GCAGAGGAAGAGCTTCAGGGGGACAGGCGGATGCAGGAAGACAGAGCGCCCTCTAGTGGCAACAGCACTAGG  
AGCGACAGCAGTGTCTGTGTGGATGACACACTGGGACAAAGTTGGGGCTGTGAAGGTCAAGGAGGAACCAAGT  
GACAGTGAAGATGCTCAGATCCAGGAAATGGAATCTGGGAGCAGGCTGCTTTTATGCAACAGCCCTTTC

FIG. 13B

109/173

CTGGAACCCACGCACACACGTGCGCTCTCTGTGCGCCAAAGCTCCGCTGGCTGCGGTGGCATGGATGGATTAGAGAAACACCGTCTCGTCTCCAGGACTCACTCTTCCCCTGCTGCTCTGTCTTTTACCTCACCCAGCAATGGACCGCCCCCTCCAGCCTGGCTCTGCAACTGGAATTGCCCTATGACCCCTTGATGCTGAAACACCAAGTCCGTTTGTGGCAATTCCACCAACCCCTGAGCATGCTGGACGAATACAGAGTATCTGGTCAAGCTGCAAGAACTGGGCTGCTAAATGAGCGAATTCAAGTCTGAAAGCCAGCCCTGGAGGAAATACAGCTTGTTCATCTTGAAATCATCACTCACTGTTGTATGGCACCAACCCCTGGACGGACAGAAAGCTGGACCCAGGATACCTCCTAGGTGATGACTCTCAAAAGTTTTTTTCCCTCATTAACCTTGTGGTGGACTTGGGGTGGACAGTGACACCATTTGGAATGAGCTACACTCGTCCGGTGTGCACGCATGGCTGTTGGCTGTGTCAATCGAGCTGGCTTCCAAAGTGGCCCTCAGGAAGCTGAAGAATGGGTTTGTGTGTGAGGCCCCCTGGCCCATCACGCTGAAGAAATCCACAGCCATGGGGTTC TGCTTTTAAATTCAGTTGCAATTACCGCCAAATACTTGAGAGACCAACTAAATATAAGCAAGATATTGATTTGTAGATCTGGATGTTCAACCATGGAAACGGTACCAGCAGGCCCTTTTATGCTGACCCAGCATCCTGTACATTTTCACCTCCATCGCTATGATGAAGGAACTTTTTTCCCTGGCAGTGGAGCCCCAAATGAGGTTTCGGTTTATTCTTTAGAGCCCCACTTTTATTGTATCTTTTCAGGTAATTGCATTGCA

FIG. 13C



110/173

FIG. 13D

tttttggttgcaagcagcagattacgcgcagaaaaaggatctcaagaagatcccttggatctttttctacggggtctgacgctcagctg  
ggaacgaaaaactcacggttaagggttttggtcatgagattatcaaaaaaggatcttcacccctagatcccttttaaatataaaatgaagtttta  
aatcaatcataaagtatatatgagtaaaacttggtctgacagttaccaatgcttaatcagtgaggcaccctatctcagcgaatctgtctatttc  
gtttatccatagttgcccgtgactcccgcgtgtagataactacgatacgggggggtttaccatctggccccaggtgctgccaatgata  
ccccggagacccacgctcacccggctccagatttatcagcaataaacacgacggcggaaaggccggagcgcagaaagtggctct  
gcaactttatccgcctccatccagctctattaattgttgccgggaaagctagagtaagtgttcgccagtttaatagtttggcgcacagttgt  
tggccattgctacaggcatcgtggtgtcacgcctcgtgttggtatggcttcaatcagctccggctcccaacgatacaaggcgaggttac  
atgatacccccatgtttgtcaaaaaaggcgttagctcttcggctctcgatcgtttgtcagaaagtaagtggccgcaggtttatcact  
catggttatggcagcacgtgcataattcttactgtcatgcccataccgtaaggatgcttttctgtgactggtgagttactcaaccaagctcat  
ctgaggaatagtgatgctggcgaccgagttgctctggccggctcaatacggggaataacccggccacatagcagaactttaaaaa  
gtgctcatcatggaaaaacgttcttcggggcgaaaaactctcaaggatcttaccgctgttggagatccagttcgatgtaaccccactcgt  
gcacccaactgatacttcagcatcttttactttaccagcgtttctgggtgagcaaaaaacaggaaggcaaaatgccgcataaaagg  
gaataaaggcgacacggaaaatgttgaatactcatacttctcttttcaataattatgaagcattttatcagggttattgctctcatgagcg  
gatacatatttgaatgtatttagaaaaataaacaataaggggttcggcgacatttccccgaaaaagtgccacctgacggccccctgt  
agcggcgcatthaagcgcgcggtgtgtgtgttacggcgagcgtgacccgtacacttggcagcgccctagcgccccggctcttt  
cgctttctccctctctcggcacgcttcggcgcttccccgtaagctctaaatcggggcatacccttttaggggttcggatttagtgc  
tttacggcacctcgacccccaaaacttgatttaggggtgtaggtttcacgtagtggggccatcgccctgatagacgggttttcggcccc  
gacgttggagctccacgcttcttaatagtgaccttgttccaaaactgggaacaacactcaacccctatctcgctctattcttttgatttataa  
gggattttggccgatttcggccctattgtttaaaaaatgagctgatttaacaaaaatttaacgcggaattttaacaaaaatattaaacggtttac  
aattt

FIG. 13E

112/173

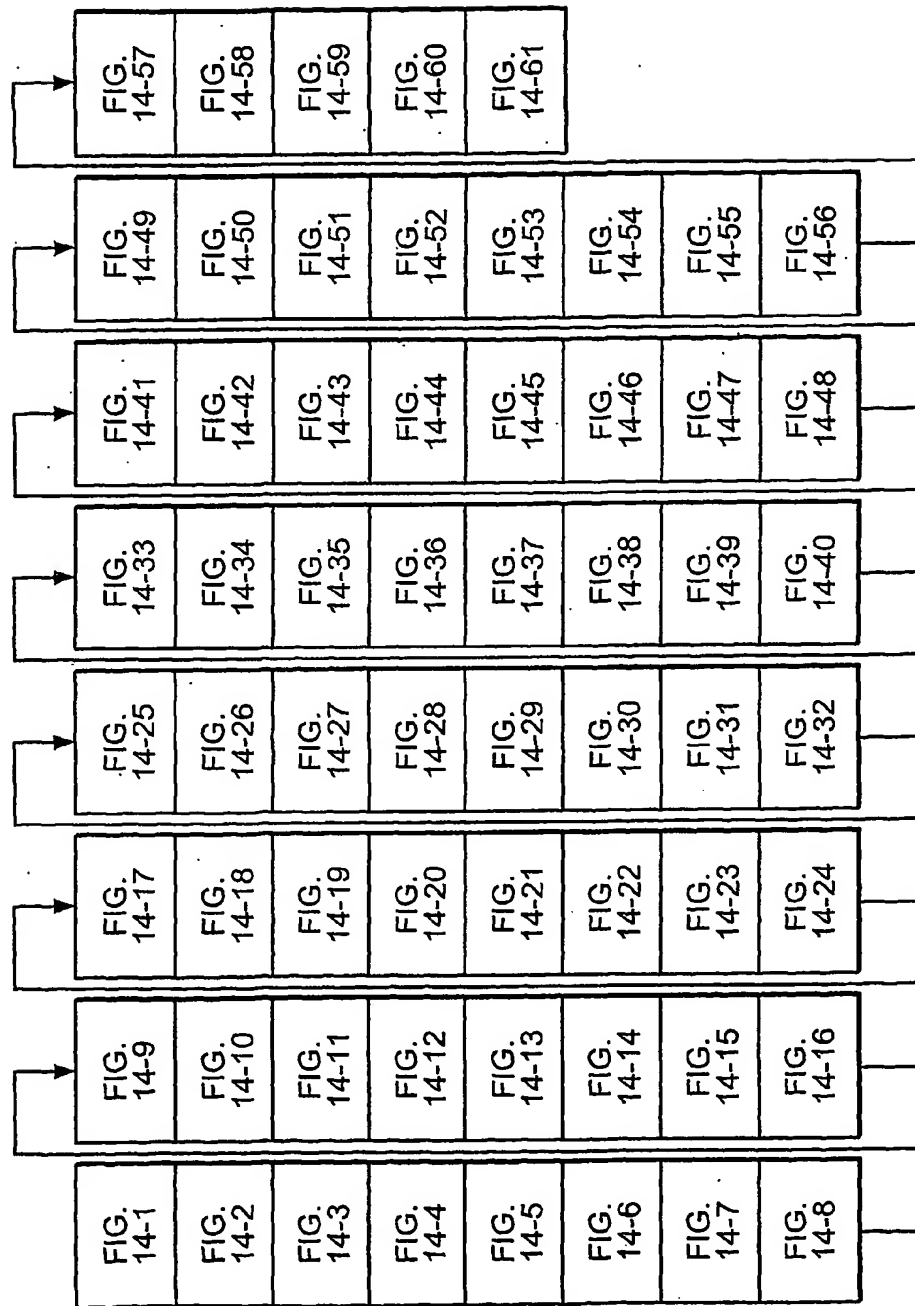


FIG. 14

pFLAG-CMV-5b-HDAC9a

7303 base pairs

Graphic map | Table by enzyme name

BstMCI		
AviII	PvuI BsiEI	EarI
BglI FspI	BsaOI	Eam1104I
cccatcgcattcaggctgcgcaactgttggaaggcgatcggtgcgggcctcttcgctattaccagctgg		MspAII
base pairs		PvuII
gggtaagcggtaagtcgcgacggttgacaacccctcccgctagccacgccccggagagcgataatgcggtcgacc		113/173
1 to 75		
Acc16I	BspCI	Ksp632I
	Bsh1285I	NspBII
	Ple19I	

FIG. 14-1

114/173

cgaaaggggatgtgctgcaaggcgattaagttgggtaacgcccagggtttcccagtcacgacggttgtaaaacg  
base pairs  
gctttcccctacacgacgttccgctaattcaacccattgagggtcccaaaagggtcagtgctgcaacatttttgc  
76 to 150

MscI

CfrI

SspI MluNI

EaeI

acggccagtgccaagctgatctaataatcaataattggccattagccataattattcattggttatatagcataaatcaa  
base pairs  
tgccggtcacgggttcgactagattagttataaaccggtaatcgggtataataagtaaccaataatcgtatttagtt  
151 to 225

EaeI

BclI

CfrI

FIG. 14-2

MscI  
 MluNI  
 SspI    EaeI    BsrDI  
 tattggctattggccattggcattacgttgatccatataataatgtacatttatattggctcatgtccaacatt  
 base pairs  
 ataaccgataaccggtaacgtatgcaacataggtagtattatacatgtgtaataataaccgagtagcaggttgtaa  
 226 to 300  
 CfrI  
 BsrGI  
 BsrDI  
 SspBI  
 Bsp1407I

115/173

HincII    VspI  
 SpeI    PshBI  
 accgccatgttgacattgattattgactagttatttaataagtaatacaattacggggtcattagttcatagcccata  
 base pairs  
 tggcgggtacaactgtaactaataactgatcaataattatcattagttaatgccccagtaatacaagtatcgggtat  
 301 to 375  
 HindII    AclNI    AsnI  
 AseI

FIG. 14-3

Hin1I  
AcyI  
HincII

BstMCI  
BglI BsaOI

tatggagtccgcgttacataaacttacggtaaatggccccgctggcgaccgcccagcgacccccgcgcttgacg  
base pairs  
atacctcaaggcgcaatgtattgaatgccattttaccggcgaccgctggcggggtcgctggggcgggcaactgc  
376 to 450

HindII  
Hsp92I  
Msp17I

Bsh1285I  
BsiEI

116/173

BsaHI  
AatII  
BbiII

BbiII  
Hin1I  
AcyI AatII

tcaatagtgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagttattacgg  
base pairs  
agttatcactgcatacaagggtatcattgcggttatccctgaaaggtaactgcagttaccacccataaaatgcc  
451 to 525

Msp17I  
BsaHI  
Hsp92I

FIG. 14-4

117/173

BbII  
HinII  
AcyI AatII  
NdeI  
BglI  
base pairs  
526 to 600

FauNDI  
Msp17I  
BsaHI  
Hsp92I

BstSNI  
SnaBI  
base pairs  
601 to 675

BsaAI  
Eco105I

**FIG. 14-5**



118/173

NcoI Bsp19I  
 StyI BstDSI  
 EcoT14I  
 atcgctattaccatgggtgatgcgggtttggcagtagcgaatggcggtgtagcggttgactcacggggattt  
 base pairs  
 tagcgataatggtagcactacgcaaaacgcgtcatgtggttaccgcacctatcgccaaactgagtgcctctaaa  
 676 to 750  
 BssT1I  
 ErhI Eco130I  
 DsaI MslI  
 BbiII  
 Hin1I  
 AccB1I  
 BshNI  
 ccaagtctccaccattgacgtcaatgggagttgttttggcaccacaaatcaacgggactttccaaaatgtcgt  
 base pairs  
 gggtcagaggtgggtaactgcagttaccctcaaaacacggtgttagttgacctgaaagggttttacagca  
 751 to 825  
 Msp17I  
 BsaHI  
 Hsp92I  
 BanI  
 Eco64I

FIG. 14-6

119/173

HincII  
 Eco24I  
 EcoICRI  
 aataacccgccccgttgacgcaaatggcggttaggcgtgtacggtgggaggtctatataagca gagctcgttta  
 base pairs  
 ttattgggggggcaactgcgtttaccgcccatccgcacatgccaccctccagatatattcgt ctcgagcaaat  
 826 to 900  
 HindII  
 Ecl136II  
 Bbv12I  
 Asphi  
 Psp124BI

SacI  
 FrioI  
 SstI  
 BanII  
 BsiHKAI  
 AcsI  
 Apol  
 HindIII  
 BstZI  
 BstMCI  
 MflI  
 Bsa29I  
 SfcI  
 EagI  
 XmaIII  
 BstYI  
 BspDI  
 BcgI  
 Eco32I  
 CciNI  
 Bsh1285I  
 BstX2I  
 BanIII  
 PstI  
 Ksp22I  
 gtgaaccgtcagaattcaagcttgcgccgcagatctatcgtatctgcaggatatcaccatgcacagtatgatcag  
 base pairs  
 cacttggcagtccttaagttcgaaacggcggtctagatagctagacgtcctatagtggtacgtgtcatactagtc  
 901 to 975  
 EcoRI  
 EaeI  
 Eco52I  
 BglII  
 Bsci  
 BspXI  
 BstSFI  
 CfrI  
 EclXI  
 BsiEI  
 BseCI  
 Bsu15I  
 EcoRV  
 NotI  
 BsaOI  
 XhoII  
 ClaI  
 Bsp106I  
 FbaI  
 Alw21I

FIG. 14-7

	FriOI	CvnI	CvnI
	Eco24I	AocI	AocI
	BpmI	Bsu36I	Bsu36I

ctcagtgatgtgaagtcagaagttcctgtggccctggagcccatctcacctttagacctaaaggacagacctcag  
base pairs

gagtcacctacacttcagtccttcaaggacacccggaccctcgggtagagtggaaatctggattcctgtctggagtc  
976 to 1050

GsuI	Eco81I	Eco81I
BanII	Bse21I	Bse21I

120/173

DsaI	DrdI	MfeI	Asp700I
------	------	------	---------

gatgatgatgcccggtggtggaccctgttgtccgtgagaagcaattgcagcaggaattacttcttatccagcagca  
base pairs

ctactactacgggaccacactgggacaaacaggcactcttcggttaacgctcgtcccttaatgaagaataggctcgtcgt  
1051 to 1125

BstDSI	MunI	XmnI
--------	------	------

FIG. 14-8

121/173

AlwNI  
gcaacaaatccagaagcagcttctgatagcagagtttcagaaacagcatgagaacttgacacggcagcaccaggc  
base pairs  
cgttgttttaggtccttcgtcgaagactatcgtctcaaaagtcttctgtcgtaactcttggaactgtgccgtcgtggtccg  
1126 to 1200

BlpI Eco57I EcoNI AlwNI  
CellII  
tcagcttcaggagcatatcaaggaaacttctagccataaaacagcaacaagaactcctagaaaaggagcagaact  
base pairs  
agtcgaagtcctcgtatagttccttgaagatcgggtattttgtcgttggtctcttgaggatcttttcctcgtcttga  
1201 to 1275  
Bsp1720I  
Bpu1102I

FIG. 14-9

122/173

BpmI  
BseRI  
ggagcagcagaggaagaacaggaaagtagagaggcatcgagagaacagcagcttcctcctctcagagggaaga  
base pairs  
cctcgtcgtctccggttcttggtcccttcattctctccgtagcgtctctgtcgaaggaggagagtcctccgtttct  
1276 to 1350  
GsuI  
EcoNI

HindIII  
tagaggacgagaaagggcagtggaagtagacagaagtaagcag aagcttcaagagttcctactgagtaaatcagc  
base pairs  
atctcctgctcttcccggtcacccgttcattgtcttcatttcgtc ttcgaagttctcaaggatgactcatttagtcg  
1351 to 1425

FIG. 14-10

Van91I  
AccB7I  
aacgaaagacactccaactaatggaaaaaatcattccgtgagccgccatcccaagctctgtacacggctgcccc  
base pairs  
ttgcttttctgtgaggttgattaccttttttagtaaggcactcggcggtagggttcgagaccatgtgccgacgggt  
1426 to 1500  
Esp1396I  
PflMI  
123/173  
ccacacatcattggatcaaagctctccaccccttagtggaacatctccatccctacaagtacacattaccaggagc  
base pairs  
gggtgtagtaaacctagtttcgagaggtggggaatcaccttgttagaggtaggatgttcatgtgtaatgggtcctcg  
1501 to 1575

FIG. 14-11

124/173

Alw21I                    BstBI  
AspHI                    Bpu14I                    FriOI  
                          Csp45I                    Eco24I  
acaagatgcaaggatgatttcccccttcgaaaaactgcctctgagcccaacttgaagggtcggtccagggttaaa  
base pairs  
tgttctacgtttctactaaaggggaagctttttgacggagactcgggttgaaacttccacgccagggtccaattt  
1576 to 1650  
BsiHKA I                    SfuI    Bsp119I                    BanII  
Bbv12I                    NspV  
                          LspI

acagaaaagtggcagagaggagaagcagcccttactcaggcgggaaggatggaaatgtgtcacttcattcaagaa  
base pairs  
tgtctttcacccgtctctctcttgtcggggaatgagtcgccttacctttacaacagtgaagtaagtctctt  
1651 to 1725

**FIG. 14-12**

125/173

Van91I	Van91I
AccB7I	AccB7I
BpmI PflMI	
gcgaatggttgaggtagacagaatcctcagtcagtagcagttctccaggctctggtcccagttcaccaacaatgg	
base pairs	
cgcttacaaactccactgtcttaggagtcagtcacgtcaagaggtccgagaccagggtcaagtggttggttacc	
1726 to 1800	
GsuI	Esp1396I
	PflMI
AlwNI	
Esp1396I	

gccaactggaagtgtactgaaaatgagacttcggttttgcacctaccctcatgccgagcaaatggtttcacacac  
base pairs  
cggttgaccttcacaatgactttactctgaagccaaaacgggggatggggagtacggctcgtttaccaaaagtgt  
1801 to 1875

**FIG. 14-13**



BsaMI  
Mva1269I  
gcaacgcattcttaattcatgaagattccatgaacctgctaagtctttatacctctccttctttgtcccaacattac  
base pairs  
cgttgcgtaagattaagtacttctaaggacttggacgattcagaaatatggagaggaacgggttgtaatg  
1876 to 1950

BspMI XcmI

BsmI RcaI  
BspHI

ErhI  
BssT1I

BstBI AcsI  
Bpu14I  
Csp45I

126/173

cttggggcttcccgcagtgccatcccagctcaatgctc gaattcactcaaagaaagcagaagtggtgagacgca  
base pairs  
gaaccccgaaagggcggtcacggtagggtaggttacgaag ctttaagtgaagtttcttttcgtcttcacactctgcgt  
1951 to 2025

Esp3I

EcoT14I

SfuI Bsp119I

BsmBI

StyI  
Eco130I

NspV ApoI  
LspI EcoRI

FIG. 14-14

127/173

MsII  
gacgcttaggcaagggtgttcctctgcctgggcagtatggaggcagcatcccggcatcttccagccaccctcatgt  
base pairs  
ctgcgaatccggttccacaaggagacggaccggtcatacctccgtcgtagggccgtagaaggtcggtgggagtaca  
2026 to 2100

PstI  
SfiI  
tactttagagggaaagccaccacagcagccaccaggctctc ctgcagcatttatttgaaagaacaaatgcg  
base pairs  
atgaaatctcccttccggtgggtgtcgtcggtgggtccgagag gacgtcgtaataataactttctgtttacgc  
2101 to 2175  
BstSFI

FIG. 14-15

128/173

Asp718I  
Acc65I  
BshNI  
BsgI

acctggcattagaggtaccacaaaattgccccgtcacagacccctgaaccgaaccaggctgcacctttgcctca  
base pairs  
tggaccgtaatctccatgggtgtttaacggggcagtgcttggggacttggcttgggtcagacgtggaaaacggagt  
2251 to 2325

BanI KpnI  
AccB1I  
Eco64I

FIG. 14-16

129/173

Bpu1102I	
Alw21I Bsp1720I	
AspHI CelII	
gagcacgttgggtcagctcattcaacagcaacaccagcaattcttggaagcagaagaataaccagcagca	
base pairs	
ctcgtgcaaccgagtcgaccagtaagttgtcggttggttaagaacctcttcgtcttcgttatgggtcgtcgt	
2326 to 2400	
BsiHKAII PvuII	
Bbv12I BlnI MspAII	
NspBII	
MflI	BstBI
XhoII	Bpu14I
	Csp45I
	Eco57I
gatccacatgaacaaactgcttctcgaaatctattgaacaaactgaagcaaccaggcagtcaccttgaggaagcaga	
base pairs	
ctagggtgtagcttgggtgacgaaagcttagataacttgtagcttcgttggtccggtcagtggaactccttcgtct	
2401 to 2475	
BstYI	SfuI Bsp119I
BstX2I	NspV
	LspI

FIG. 14-17

130/173

EarI  
 Eam1104I  
 Asp700I  
 Bbv16II  
 BbsI Bsp143II  
 ggaagagcttcaggggaccaggcgatgcaggaagacagagcgccctctagtggcaacagcactaggagcgacag  
 base pairs  
 ccttcgaagtcctccctgggtccgctacgtccttctgtctcggggagatcacggttgtcgtgatccctcgctgtc  
 2476 to 2550  
 XmnI Eco57I  
 Ksp632I  
 SspI  
 BpiI HaeII  
 BpuAI BstH2I

BcgI  
 cagtgtgtgtggatgacacactgggacaagttggggctgtgaaggtcaaggagggaaccagtggaacagtgatga  
 base pairs  
 gtcacgaacacacctactgtgtgacctgttcaaccccgacacttcaggttcctccttggtcacctgtcactact  
 2551 to 2625

FIG. 14-18

MflI Van91I  
 XhoII AccB7I  
 agatgctcagatccaggaaatggaatctggggagcaggctgcttttatgcaacagcctttcctggaacccacgca  
 base pairs  
 tctacgagctcaggtcctttaccttagacccctcgccgacgaaatacgttgctggaaaggaccttgggtgcgt  
 2626 to 2700  
 BstYI Esp1396I  
 BstX2I PflMI

131/173

PmaCI  
 PmlI  
 AflIII  
 NspBII  
 Esp3I  
 cacacgtgcgctctctgtgcgccaaagctccgctggctgcggttgccatggatggattagagaaacacccgtctcgt  
 base pairs  
 gtgtgcacgcgagagacacgcggttcgaggcaccgacgccaaccgtacctaatctcttgtggcagagca  
 2701 to 2775  
 MslI Eco72I  
 MspA1I  
 BsmBI  
 BsaAI  
 BbrPI

FIG. 14-19

EarI  
 BpmI      BpmI  
 Eam1104I      BsrDI  
 ctccaggactcactcttccctgctgctctgtttacctcaccagcaatggaccgccccctccagcctggctc  
 base pairs  
 gaggtcctgagtgaaggacgacggagacaaaatggagtgggtcgttacctggcgggggaggtcggaccgag  
 2776 to 2850  
 GsuI      Ksp632I      GsuI  
 132/173  
 XcmI  
 tgcaactggaattgcctatgaccccttgatgctgaacaccagtgcggttggtggcaattccaccaccacctga  
 base pairs  
 acgttgaccttaacggatactggggaactacgactttgtggtcacgcaaacaccgttaagggtgggtgggact  
 2851 to 2925

FIG. 14-20

133/173

<p>SphI BbuI</p> <p>gcattgctggacgaatacacagagtattcgtcacgactgcaagaaactgggctgctaaataaatgtgagc gaatttca base pairs</p> <p>cgtacgacctgcttatgtctcatagaccagtgctgacgttctttgacccgacgatttatttacctcg cttaagt 2926 to 3000</p> <p>PaeI NspI</p>	<p>AcsI ApoI</p>	<p>EcoRI</p>
<p>BpmI</p> <p>aggtcgaaaagccagcctggaggaaatacacagcttggtcattctgaacatcactcactgttgatggcaccaacc base pairs</p> <p>tccagcttttcggtcggacctcctttatgtcgaacaagtaagactttagtgagtgaacaacataccgtggttggg 3001 to 3075</p> <p>GsuI</p>	<p>AccB1I BshNI</p>	<p>BanI Eco64I</p>

FIG. 14-21



134/173

ErhI  
StyI Eco130I  
EcoT14I  
BstXI AlwNI  
cctggacggacagaagctggacccaggatactcctaggtgatgactctcaaaagtttttccctcattaccttg  
base pairs  
ggacctgcctgtcttcgacctggggtccctatgaggatccactactgagagttttcaaaaaaggagtaatggaac  
3076 to 3150

BssT1I  
AvtII  
BlnI

BsaWI BsgI  
tggaggacttggggtggacagtgacaccatttggaatgagctacactcgctccggtgctgcacgcacatggctgttg  
base pairs  
accacctgaacccacacctgtcactgtggtaaaccttactcgatgtgagcaggccacgacgtgcgtaccgacaacc  
3151 to 3225

FIG. 14-22

CvnI				
AocI				
Bsu36I	Eco57I			
		DraII	EaeI	
ctgtgtcatcgagctggcttccaaagtggcctcaggagagctgaagaatgggtttgctgtgtgtagggccccctgg				
base pairs				
gacacagtagctcgaccgaagggtttcacccggagtcctctcgacttcttaccacaacgacacactccgggggacc				
3226 to 3300				
Eco81I				Eco0109I
Bse21I				

135/173

MscI	ErhI	Eco130I		
	BssT1I	BstXI		
	Eco57I	MslI	DsaI	
ccatcacgctgaagaatccacagccatgggggttctgcttttttaattcagttgcaattaccgccaatacttgag				
base pairs				
ggtagtgcgacttcttaggtgctcggtaccccaagacgaaaaaattaaagtcaacggttaatggcggttttatgaactc				
3301 to 3375				
MluNI	EcoT14I			
BalI	StyI	BstDSI		
	NcoI	Bsp19I		

FIG. 14-23

Eco147I  
 BsaI  
 agaccaactaaataaagcaagataattgattgtagatctggatgttcaccatggaaacgggtaccacagcaggcctt  
 base pairs  
 tctgggttgatttatattcgttctataactaactatctagacctacaagtgggtacctttgccatgggtcgtccggaa  
 3376 to 3450  
 Eco31I  
 BstX2I  
 BstYI  
 XhoII  
 NcoI Bsp19I Asp718I SseBI  
 StyI BstDSI AccB1I  
 EcoT14I BshNI StuI  
 BglII  
 MflI  
 BssT1I BanI KpnI AatI  
 ErhI Eco130I Eco64I Pme55I  
 DsaI Acc65I  
 136/173

SspBI  
 Bsp1407I MslI Asp700I  
 ttatgctgacccagcatcctgtacatttccactccatcgctatgatgaagggaacttttccctggcagtgaggc  
 base pairs  
 aatacgaactggggtcgtaggacatgtaaagtgaaggtagcgatactacttcccttgaaaaaggaccgtcacctcg  
 3451 to 3525  
 BsrGI  
 XmnI

FIG. 14-24

FriOI	FriOI	BstYI
Eco24I	Eco24I	XhoII
		BsrDI
cccaaatgaggttcggttatttcttagagccccacttttatttgatatctttcaggtaattgcattgca ggatc		
base pairs		
gggtttactccaagccaataaaagaatctcggggtgaaaaataaacatagaaagtcatttaacgtaacgt cctag		
3526 to 3600		
BanII	BanII	BamHI
		BstI
		MflI

137/173

Acc65I	AvaI BcoI
BanI Eco64I	MflI Eco88I PspALI
BstX2I Asp718I	XhoII Cfr9I SmaI MslI
cggtaaccagattacaaggacgacgatgacaaagtagat cccgggtggcatccctgtgacccccctcccagtgccctct	
base pairs	
gccatgggtctaattgttcctgctgctactgttcatcta gggcccaccgtaggggacactggggaggggtcacgggaga	
3601 to 3675	
BshNI	BstYI Ama87I
BsaWI KpnI	BstX2I BsoBI
AccB1I	XmaI PspAI

FIG. 14-25

Eco130I  
 StyI  
 EcoT14I  
 GsuI  
 MslI  
 cctggccttggaagtgcactccagtgcccaaccagccttgctcctaataaaattaagttgcatcattttgtctga  
 base pairs  
 ggaacggaaccttcaacggtgaggtcacgggtggtcggaacaggattattttaattcaacgtagtaaaacagact  
 3676 to 3750  
 BpmI  
 BssT1I  
 ErhI  
 138/173  
 Eco24I  
 SfcI  
 Bbv16II  
 BbsI  
 PspOMI  
 FrioI  
 Eco1105I  
 AspEI  
 SspI  
 ctagggtgtcctctataattatgggtggaggggggtgtatggagcaagggggcccaagttgggaagacaacct  
 base pairs  
 gatccacaggagatatataataacccacacctccccaccataacctcggtccccgggtcaacccttctgttgga  
 3751 to 3825  
 Bsp120I  
 EcoO109I  
 BpiI  
 BpuAI  
 BstSFI  
 AhdI  
 ApaI

FIG. 14-26

DraII	BpmI BsgI		
gtagggcctgcgggtctattcgggaaccaagctggagtgagtgggcacaaatcttggctcactgcaatctccgcc			
base pairs			
catccgggacgccccagataagcccttggttcgacctcacgtcacctgttagaaccgagtgacgttagaggcgg			
3826 to 3900	GsuI		
EcoO109I			
			139/173
	BcoI	NspI	BlpI
	Ama87I	PaeI	Mph1103I
		Ppu10I	EcoT22I
	BcgI		
tcctgggttcaagcgattctcctgcctcagcctcccaggtgtgttgggattccaggcatgcatgaccagggtcagc			
base pairs			
aggacccaagtctcgtaagaggacggagtcggagggtcaacaaccctaagggtccgtactgtggtccgagtcg			
3901 to 3975	Eco88I	BbuI Zsp2I	CellI
	BsoBI	SphI	Bsp172
		NsiI	Bpu1I

FIG. 14-27

140/173

		MscI	
		MluNI	
	Esp3I	EaeI	BsaI
taatttttggttttttggttagagacggggtttccaccatattggccagggtgtctccaactcctaattctcaggtg			
base pairs			
attaaaaacaaaaaaccatctctgtccccaaagtgggtataaccgggtccgaccagaggttgaggattagagtcac			
3976 to 4050			
	BsmBI	CfrI	Eco31I
0I		BalI	
02I			
	Eco130I		
	StyI		
	EcoT14I	BstXI	
atctaccaccttgccctcccaaatgtctgggttacaggcgtgaaccactgctcccttccctgtccttctgatt			
base pairs			
tagatgggtggaaccggagggtttaacgaccctaattgtccgcacttggtagcaggggaaggacaggaagactaa			
4051 to 4125			
	BstT1I		
	ErhI		

FIG. 14-28

BbiII	NcoI	Eco130I	BsrFI	PflMI
HinII	StyI	DsaI	AgeI	Bse118I
AcyI	EcoT14I	BsaWI	AccB7I	
DraI				
ttaaaataataaccagcaggagacgtccagacacagcataggctacctgccatgccccaaaccggtgggacat				
base pairs				
aattttattgatattggtcgtcctcctgcaggctctgtgtcgtatccgatggacggtaccgggttgccaccctgta				
4126 to 4200				
Msp17I	BssAI	Esp1396I		
BsaHI	ErhI	BstDSI	PinAI	Van91I
Hsp92I	BspMI	Bsp19I	Cfr10I	
				141/173
ttgagttgcttggtggcactgtcctctcatgcgttggtgggtccactcagtagatgcctgttgaattgggtacgcgg				
base pairs				
aactcaacgaacacctgacaggagagtagcgcaacccaggtagtcattacggacaacttaaccccatgcgcc				
4201 to 4275				
				EaeI
				CfrI

FIG. 14-29



142/173

AlwNI

ccagcttctgtggaatgtgtgtcagttaggtgtggaaagtccccagggtccccagcaggcagaagtatgcaaaag  
base pairs  
ggtcgaagacacaccttacacacagtcfaatccccacacctttcaggggtccgaggggtcgtccgtcttcatacgttttc  
4276 to 4350

NspI

PaeI Mph1103I

Ppu10I EcoT22I SexAI

catgcatctcaattagtcagcaaccagggtgtggaaaaagtccccagggtccccagcaggcagaagtatgcaaaagca  
base pairs  
gtacgtagagttaatcagtcgttggtccacacaccttttcaggggtccgaggggtcgtccgtcttcatacgttttcgt  
4351 to 4425

BbuI Zsp2I

SphI

NsiI

FIG. 14-30

143/173

NspI  
PaeI Mph1103I  
Ppu10I EcoT22I  
tgcatctcaattagtcagcaaccatagtcgcccccctaactccgcccataccgcccctaactccgcccagttccg  
base pairs  
acgtagagtttaatcagtcggttggtatcagggcggggattgagggcggggattgagggcgggtcaaggc  
4426 to 4500  
BbuI Zsp2I  
SphI  
NsiI

NcoI Bsp19I  
StyI BstDSI  
EcoT14I  
BglI  
cccatctccgcccctggctgactaatTTTTTTTatttatgcagagccgagccgctcggcctctgagctat  
base pairs  
gggtaagagcggggtaccgactgattaaaaaaaaataacgtctccggctccggcggagccggagactcgata  
4501 to 4575  
BssT1I  
ErhI Eco130I  
DsaI  
SfiI

FIG. 14-31

144/173

SseBI AvrII Ama87I  
 Eco147I BlnI Eco88I BseRI  
 StuI BssTII AvaI BsoBI  
 BseRI  
 tccagaagtagtgaggaggcttttttgaggccctaggcttttgcaaaaagctc ctcgaggaaactgaaaaaccaga  
 base pairs  
 aggtcttcactcctccgaaaaaacctccggatccgaaaacgttttttcgag gagtccttgacttttttggtct  
 4576 to 4650

AatI StyI XhoI BcoI  
 Pme55I ErhI Sfr274I  
 EcoT14I Eco130I Paer7I

SfcI ApoI  
 aagttaattccctatagtgagtcgtatttaaattcgtaatcatggtcatagtgttctctgtgtgaaattgtttatc  
 base pairs  
 ttcaattaagggatatactcagcataaatttaagcattagtagaccagtagatcgacaaaaggacacactttaacaatag  
 4651 to 4725

BstSFI AcsI

FIG. 14-32

AccBSI  
BsrBI

cgctcacaattccacacaacatacagagcgggaagcataaagtgtaaagcctggggtgcctaataatgagtgagctaac  
base pairs  
gcgagtgttaagggtgtgtatgctcggccttcgtatttcacatttcggagccccacggattactcactcgattg  
4726 to 4800  
BstD102I

AccBII  
BshNI

VspI  
PshBI

tcacattaattgcgctgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaatgaatcg  
base pairs  
agtgtaatcaacgcaacgcgagtgacgggcgaaaaggtcagccctttggacagcacggtcgacgtaattacttagc  
4801 to 4875  
AsnI  
AseI

VspI  
MspAII  
PvuII PshBI  
EaeI

NspBII  
AsnI  
AseI

NspBII  
CfrI

145/173

FIG. 14-33

146/173

Eam1104I  
 BstH2I  
 Bsp143II  
 gccaacgcgcgggagagcggttgcgtattgggcgctcttcgcgttcctcgctcactgactcgctgcgctcgg  
 base pairs  
 cggttgcgcgccccctctccgccaaacgcataaacccgcgagaaaggcagagtgactgagcgacgcgagcc  
 4876 to 4950

HaeII EarI  
 SapI  
 Ksp632I

BstMCI  
 BsaOI  
 AccBSI  
 BsrBI  
 tcggttcggctgcgcgagcggtatcagctcactcaaaaggcggttaatacggttatccacagaatcaggggataacg  
 base pairs  
 agcaagccgacgcgcgctcgccatagtcgagtgagttccgcgcattatgcccaatagggtgtcttagtccccctattgc  
 4951 to 5025  
 Bsh1285I  
 BsiEI  
 BstD102I

FIG. 14-34

147/173

NspI  
BspLU11I  
caggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcttgctggcgtttttcc  
base pairs  
gtcctttcttgtaactcgttttcggtcgttttcggtccttggcatttttcggcgcaaacgacgcgcaaaaagg  
5026 to 5100  
AflIII

DrdI  
ataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaacccgacaggactat  
base pairs  
tatccgagcgggggactgctcgtagtgttttagctcgaggtcagtcctccaccgctttgggctgtcctgata  
5101 to 5175

FIG. 14-35

148/173

BsaWI  
BsiI  
aaagataccaggcggtttccccctgggaagctccctcgtgcgtctctcctgttcgacccctgccgcttaccggatacc  
base pairs  
tttctatggtccgcaaaaggggaccttcgaggagcacgcgagaggaagctgggacggcggaatggcctatgg  
5176 to 5250

BssSI

BstH2I  
Bsp143II  
SfcI  
tgtccgcctttctcccttcgggaagcgtggcgctttctcaatgctcacgctgtaggtatctcagttcgggtgtagg  
base pairs  
acaggcggaaagagggaagcccttcgcaccgcgaaagagttacgagtgcgacatccatagagtcaagccacatcc  
5251 to 5325

HaeII  
BstSFI

FIG. 14-36

149/173

BsiHKAI                      NspBII  
                                  BstMCI  
                                  BsaOI                      BsaWI  
 VneI Bbv12I  
 tcgttcgctccaaagctgggctgtgtgcacgaaccccccggttcagcccgaccgctgcgcccttatccggtaactatc  
 base pairs  
 agcaagcgagggttcgacccgacacacgtgcttggggggcaagtcgggctggcgacgcggaataggccattgatag  
 5326 to 5400  
 ApaLI                      Bsh1285I  
                                  AspHI                      BsiEI  
                                  Alw21I                      MspA1I

                                 AlwNI  
 gtcttgagtccaaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcga  
 base pairs  
 cagaactcaggttggggccattctgtgctgaatagcggtgaccgtcgtcggtgaccattgtcctaatacgtctcgct  
 5401 to 5475

FIG. 14-37



150/173

SfcI

ggtatgtaggcgtgctacagagttcttgaagtggcctaactacggctacactagaagaacagtatattggta  
base pairs  
ccatacatccgccacgatgtctcaagaacttcaccaccggattgatgccgatgtgatcttcttgcataaaacat  
5476 to 5550

BstSFI

Eco57I

tctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaaaacaccacgctg  
base pairs  
agacgcgagacgacttcggtcaatggaagcccttttctcaaccatcgagaaactaggccgtttgttttgggtggcgac  
5551 to 5625

NspBII

MspAII

FIG. 14-38

	MflI	MflI
	XhoII	XhoII
gtagcgggtgggtttttgtgttgcaagcagcagattacgcgcagagaaaaaaggatctcaagaagatccttttgatct		
base pairs		
catcgccaccaaaaaaacaacgttcgtcgtctaatacgcgctcttttttcctagagttcttctaggaactaga		
5626 to 5700	BstYI	BstYI
	BstX2I	BstX2I

151/173

	RcaI	MflI
		XhoII
tttctacggggtctgacgctcagtggaacgaaaaactcacgttaagggttttggtcatgagattatcaaaaagga		
base pairs		
aaagatgccccagactgcgagtcaccttgcttttgagtgaattccctaaaaaccagtactctaatagttttttcct		
5701 to 5775	BspHI	BstYI
		BstX2I

FIG. 14-39

152/173

MfII                      XhoII                      DraI                      DraI  
tcttcacctagatcccttttaaatataaaatgaagttttaaatcaatctaaagtatatatgagtaaaacttggtctg  
base pairs  
agaagtggatcttaggaaaatttaatttttacttcaaaatttagttagatttcatatatactcatttgaaccagac  
5776 to 5850  
BstYI  
BstX2I

AccB1I  
BshNI  
acagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgac  
base pairs  
tgtcaatgggttacgaattagtcactccgtggatagagtcgctagacagataaaagcaagtaggtatcaacggactg  
5851 to 5925  
BanI  
Eco64I

FIG. 14-40

153/173

Eam1105I  
AspEI  
BsrDI  
tccccgctcgtgtagataactacgatacggagggttaccatctgccccagtgtgctgcaatgataccgcgagacc  
base pairs  
aggggcagcacatctattgatgctatgccctcccgaatggtagaccgggtcacgacgttactatggcgctctggg  
5926 to 6000  
EclHKI  
AhdI

Cfr10I  
BsaI BssAI BpmI  
BglI  
cacgctcacccggctccagatttatcagcaataaaccagccagccggaaggcccgagcgagaagtggctcctgcaa  
base pairs  
gtgcgagtggccgaggctctaaatagtcgttatttggtcggctcccggtcgcgtcttcaccaggacgtt  
6001 to 6075  
Eco31I BsrFI GsuI  
Bse118I

FIG. 14-41

154/173

VspI  
PshBI  
ctttatccgcctccatccagtcctatttaattgttgcgggaagctagagtaagtagttcgccagttaatagtttgc  
base pairs  
gaaataggcgggtaggtcagataaattaacaacggcccttcgatctcattcatcaagcgggtcaattatcaaacg  
6076 to 6150

AsnI  
AseI

AviII  
FspI  
gcaacgttggtgccattgctacaggcatcgtggtgcacgctcgtcgttgggtatggcttcattcagctccggtt  
base pairs  
cgttgcaacaacggtaacgatgtccgtagcaccacagtcggagcagcaaacccataccggaagtaagtcgaggccaa  
6151 to 6225  
Acc16I  
BsrDI  
Psp1406I  
BstSFI  
SfcI  
MslI  
BsaWI

FIG. 14-42

BsiEI  
PvuI  
BstMCI  
BsaOI

cccaacgatcaaggcgaggttacatgatcccccatgtgtgcaaaaaagcggttagctccttcggtcctccgatacg  
base pairs  
gggttgctagtccgctcaatgtactaggggtacaacacgttttttcgccaatcgaggaagccaggaggtacg  
6226 to 6300

BspCI  
Bsh1285I  
Ple19I

155/173

MslI

EaeI

ttgtcagaagtaagtggccgcagtggtatcactcatgggtatggcagcactgcataattcttactgtcatgc  
base pairs  
aacagtcttcattcaaccggcggtcacaaatagtgagtaccaataaccgtcgtgacgtattaaagagaatgacagtagc  
6301 to 6375

CfrI

FIG. 14-43

Acc113I	BstMCI
Eco255I	BsaOI
catccgtaagatgctttctgtgactgggtgactcaaccaagtcattctgagaatagtgtatcgggcgaccga	
base pairs	
gtaggcatttctacgaaaagacactgaccactcatgagttgggttcagtaagactcttaccatacgccgctggct	
6376 to 6450	
ScaI	Bsh1285I
	BsiEI
	156/173
BbiI	Alw21I
Hin1I	AspHI
BcgI	DraI
gttgctcttgcccggtcaatacgggataataccgcgccacatagcagaactttaaaagtgtcatcattggaa	
base pairs	
caacgagaacggccgcagttatgccctattatggcgcggtgtatcgctcttgaaattttcacgagtagtaacctt	
6451 to 6525	
Msp17I	BsiHKA1
BsaHI	Bbv12I
Hsp92I	

FIG. 14-44

157/173

XmnI	MflI	MflI	BssSI
Psp1406I	XhoII	NspBII XhoII	Alw44I
aacgttcttcggggcgaaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcac			VneI
base pairs			
ttgcaagaagccccgctttttgagagttcctagaatggcgacaactctaggtcaagctacattgggtgagcaagtg			
6526 to 6600	BstYI	MspAII BstYI	ApalI
Asp700I	BstX2I	BstX2I	BsiI
			AspHI
Bbv12I	Eco57I		
BsiHKAI			
ccaaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaa			
base pairs			
ggttgactagaagtcgtagaaaatgaaagtggtcgcaaaagaccactcgttttgccttcggttttacggcggtt			
6601 to 6675			
Alw21I			

FIG. 14-45



158/173

MsII  
EamI  
Eam1104I SspI  
aaaagggaataagggcgacacgggaatgttgaatactcatactcttcctttttccaatatattgaagcatttattc  
base pairs  
ttttcccttattcccgctgtgcctttacaacttatgagtatgagaagggaagttataataacttcgtaaatag  
6676 to 6750  
Ksp632I

RcaI  
AccBSI  
BsrBI  
agggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaataaggggttccgcgcacat  
base pairs  
tcccaataacagagtagtgcgcctatgtataaaacttacataaatcttttatttatttcccccaaggcggtgta  
6751 to 6825  
BspHI BstD102I

FIG. 14-46

SfiI

ttccccgaaaagtgccacctgacgcgccctgtagcggcgcatgaagcgcggcggtgtggtggttacgcgcagcg  
 base pairs  
 aagggcgtttttcacgggtggactgcgcgggacatcgccgcgtaattcgcgccgccacacccaatgcgcgtcgc  
 6826 to 6900

BstSFI

159/173

<p style="text-align: center;">AccBSI</p> <p>BstH2I   HaeII   BstD102I</p> <p>Bsp143II   BsrBI</p> <p>tgaccgctacacttgccagcgccctagcgcgcgccttctcgcttctcccttcttctcgccacgttcgcg              base pairs              actggcgatgtgaacggtcgcgggacgcgggcgaggaagcgaagaggaaagagcggtgcaagcggc              6901 to 6975</p>	<p style="text-align: center;">BsrFI</p> <p>BssAI</p> <p>MroNI</p> <p style="text-align: center;">NgoAIV</p> <p>NgomI</p> <p>Bse118I</p>
---	--

HaeII   Bsp143II

BstH2I

FIG. 14-47

160/173

NaeI  
gctttcccgctcaagctctaaatcggggcatccctttaggggtccgatttagtgctttacgggcacctcgacccca  
base pairs  
cgaaggggcagtttcgagatttagccccgtagggaaatcccaaggctaaatcacgaaatgccgtggagctgggggt  
6976 to 7050

AccB1I  
BshNI

BanI  
Eco64I

Cfr10I

BsaAI  
aaaaacttgattagggatggttcacgtagtgggccatcgccctgatagacgggttttcgccctttgacgttgg  
base pairs  
tttttgaactaatcccactaccaagtgcaccccgtagcgggactatctgccccaaagcgggaaactgcaacc  
7051 to 7125

DraIII

DrdI

FIG. 14-48

agtccacgttctttaatagtggaactcttgttccaactggaaacacactcaaccctatctcggctctattcttttg  
 base pairs  
 tcaggtgcaagaattatcacctgagaacaaaggtttgaccttggtgtgagttgggatatagagccagataagaaaaac  
 7126 to 7200

atttataagggaatttgcggatttcggcctattgggttaaaaatgagctgatttaacaaaaattttaacgcgaatt  
 base pairs  
 taaatattccctaaaacggctaagccggataaccaatttttactcgactaaaattgtttttaaattgcgcttaa  
 7201 to 7275

161/173

ApoI ApoI

Acsi Acsi

SspI Psp1406I  
 ttaacaaaatattaaacggtttacaattt base pairs  
 aattgttttataatttgcaaatgtttaa 7276 to 7303

FIG. 14-49

Table by Enzyme Name

Enzyme name	No. cuts	Positions of sites	Recognition sequence
AatI	2	3446 4606	agg/cct <u>More info</u>
AatII	5	451 504 587 773 4154	gacgt/c <u>More info</u>
Acc113I	1	6408	agt/act <u>More info</u>
Acc16I	2	21 6150	tgc/gca <u>More info</u>
Acc65I	3	2264 3434 3602	g/ gtacc <u>More info</u>
AccB1I	8	791 2264 3065 3434 3602 4779	g/ gyrcc <u>More info</u>
		5876 7036	
AccB7I	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg <u>More info</u>
AccBSI	4	4730 4971 6772 6936	gagcgg <u>More info</u>
Ac1NI	1	326	a/ ctagt <u>More info</u>
AcSI	7	912 1990 2244 2994 4679 7260	r/ aatty <u>More info</u>
		7271	
AcyI	6	448 501 584 770 4151 6465	gr/cgyc <u>More info</u>
AflIII	2	2702 5035	a/ crygt <u>More info</u>
AgeI	1	4188	a/ ccggt <u>More info</u>
AhdI	2	3754 5928	gacnnn/nngtc <u>More info</u>
Alw21I	6	894 1576 2330 5353 6514 6599	gwgwc/c <u>More info</u>
Alw44I	2	5349 6595	g/ tgcac <u>More info</u>

162/173

FIG. 14-50

163/173

AlwNI	6	1147	1273	1775	3091	4282	5451	cagnnn/ctg	More info
Ama87I	3	3638	3934	4629				c/ ycgrg	More info
AocI	3	1034	1046	3256				cc/ tnagg	More info
Apal	1	3806						gggcc/c	More info
ApalI	2	5349	6595					g/ tgcac	More info
ApOI	7	912	1990	2244	2994	4679	7260	r/ aatty	More info
		7271							
AseI	4	334	4806	4865	6100			at/ taat	More info
AsnI	4	334	4806	4865	6100			at/ taat	More info
Asp700I	4	1107	2481	3506	6527			gaann/nnttc	More info
Asp718I	3	2264	3434	3602				g/ gtacc	More info
AspEI	2	3754	5928					gacnnn/nngtc	More info
AspHI	6	894	1576	2330	5353	6514	6599	gwgwc/c	More info
AvaI	3	3638	3934	4629				c/ ycgrg	More info
AviII	2	21	6150					tgc/gca	More info
AvrII	2	3109	4607					c/ ctagg	More info
BalI	4	184	238	3300	4018			tgg/cca	More info
BamHI	1	3596						g/ gatcc	More info
BanI	8	791	2264	3065	3434	3602	4779	g/ gyrcc	More info
		5876	7036						
BanII	6	894	1017	1623	3526	3558	3806	grgcy/c	More info
BanIII	1	939						at/ cgat	More info
BbiII	6	448	501	584	770	4151	6465	gr/cgyc	More info

FIG. 14-51

BbrPI	1	2705		cac/gtg	<u>More info</u>
BbsI	2	2512 3820		gaagac	<u>More info</u>
BbuI	4	2930 3959 4354 4427		gcatg/c	<u>More info</u>
Bbv12I	6	894 1576 2330 5353 6514 6599		gwgw/c	<u>More info</u>
Bbv16II	2	2512 3820		gaagac	<u>More info</u>
BcgI	4	941 2556 3925 6455		cgannnnntgc	<u>More info</u>
BclI	1	969		t/ gatca	<u>More info</u>
BcoI	3	3638 3934 4629		c/ ycgrg	<u>More info</u>
BglI	5	14 417 538 4560 6048		gccnnnn/nggc	<u>More info</u>
BglII	2	932 3409		a/ gatct	<u>More info</u>
BlnI	2	3109 4607		c/ ctagg	<u>More info</u>
BlpI	3	1200 2337 3970		gc/tnagc	<u>More info</u>
BpiI	2	2512 3820		gaagac	<u>More info</u>
BpmI	9	1015 1279 1772 2781 2842 3022		ctggag	<u>More info</u>
		3701 3863 6018			
Bpu1102I	3	1200 2337 3970		gc/tnagc	<u>More info</u>
Bpu14I	3	1603 1988 2423		tt/cgaa	<u>More info</u>
BpuAI	2	2512 3820		gaagac	<u>More info</u>
Bsa29I	1	939		at/ cgat	<u>More info</u>
BsaAI	3	666 2705 7077		yac/gtr	<u>More info</u>
BsaHI	6	448 501 584 770 4151 6465		gr/cgyc	<u>More info</u>

164/173

FIG. 14-52

165/173

BsaI	3	3380	4031	6000	ggtctc	<u>More info</u>
BsaMI	1	1886			gaatgc	<u>More info</u>
BsaOI	7	42	424	928	cgry/cg	<u>More info</u>
BsaWI	6	3200	3599	4188	w/ ccggw	<u>More info</u>
BscI	1	939			at/ cgat	<u>More info</u>
BseI118I	3	4188	6008	6972	r/ ccgy	<u>More info</u>
Bse21I	3	1034	1046	3256	cc/ tnagg	<u>More info</u>
BseCI	1	939			at/ cgat	<u>More info</u>
BseRI	4	1337	1671	4593	gaggag	<u>More info</u>
BsgI	3	2315	3212	3868	gtgcag	<u>More info</u>
Bsh1285I	7	42	424	928	cgry/cg	<u>More info</u>
BshNI	8	791	2264	3065	g/ gyrcc	<u>More info</u>
		5876	7036			
BsiEI	7	42	424	928	cgry/cg	<u>More info</u>
BsiHKAI	6	894	1576	2330	gwgw/c	<u>More info</u>
BsiI	2	5213	6597		ctcgtg	<u>More info</u>
BsmBI	3	2023	2773	4001	cgtctc	<u>More info</u>
BsmI	1	1886			gaatgc	<u>More info</u>
Bsobi	3	3638	3934	4629	c/ ycgrg	<u>More info</u>
Bsp106I	1	939			at/ cgat	<u>More info</u>
Bsp119I	3	1603	1988	2423	tt/cgaa	<u>More info</u>
Bsp120I	1	3802			g/ ggccc	<u>More info</u>
Bsp1407I	2	270	3471		t/ gtaca	<u>More info</u>

FIG. 14-53



166/173

Bsp143II	5	2519	4913	5283	6922	6930	rgcgc/y	More info
Bsp1720I	3	1200	2337	3970			gc/tnagc	More info
Bsp19I	5	686	3324	3424	4178	4514	c/ catgg	More info
BspCI	2	42	6298				cgat/cg	More info
BspDI	1	939					at/ cgat	More info
BspHI	3	1891	5755	6763			t/ catga	More info
BspLU11I	1	5035					a/ catgt	More info
BspMI	2	1913	4178				acctgc	More info
BspXI	1	939					at/ cgat	More info
BsrBI	4	4730	4971	6772	6936		gagcgg	More info
BsrDI	5	245	2827	3594	5987	6169	gcaatg	More info
BsrFI	3	4188	6008	6972			r/ ccggy	More info
BsrGI	2	270	3471				t/ gtaca	More info
BssAI	3	4188	6008	6972			r/ ccggy	More info
BssSI	2	5213	6597				ctcgtg	More info
BsST1I	11	686	1950	2226	3109	3324 3424	c/ cwwgg	More info
		3681	4060	4178	4514	4607		
BstBI	3	1603	1988	2423			tt/cgaa	More info
BstD102I	4	4730	4971	6772	6936		gagcgg	More info
BstDSI	6	686	1062	3324	3424	4178 4514	c/ crygg	More info
BstH2I	5	2519	4913	5283	6922	6930	rgcgc/y	More info
BstI	1	3596					g/ gatcc	More info

FIG. 14-54

BstMCI	7	42 424 928 4951 5375 6298 6447	cgry/cg	<u>More info</u>
BstSFI	8	944 2144 3824 4662 5300 5491	c/ tryag	<u>More info</u>
		6169 6854		
BstSNI	1	666	tac/gta	<u>More info</u>
BstX2I	12	932 2400 2634 3409 3596 3634	r/ gatcy	<u>More info</u>
		5676 5687 5773 5785 6553 6570		
BstXI	3	3076 3325 4077	ccannnnn/ntgg	<u>More info</u>
BstYI	12	932 2400 2634 3409 3596 3634	r/ gatcy	<u>More info</u>
		5676 5687 5773 5785 6553 6570		
BstZI	1	925	c/ ggccg	<u>More info</u>
Bsu15I	1	939	at/ cgat	<u>More info</u>
Bsu36I	3	1034 1046 3256	cc/ tnagg	<u>More info</u>
CciNI	1	925	gc/ggccgc	<u>More info</u>
CelII	3	1200 2337 3970	gc/tnagc	<u>More info</u>
Cfr10I	3	4188 6008 6972	r/ ccggy	<u>More info</u>
Cfr9I	1	3638	c/ ccggg	<u>More info</u>
CfrI	9	152 182 236 925 3298 4016 4273	y/ ggccr	<u>More info</u>
		4874 6316		
Clai	1	939	at/ cgat	<u>More info</u>
Csp45I	3	1603 1988 2423	tt/cgaa	<u>More info</u>
CvnI	3	1034 1046 3256	cc/ tnagg	<u>More info</u>

167/173

FIG. 14-55

DraI	4	4127	5794	5813	6505	ttt/aaa	<u>More info</u>
DraII	3	3291	3802	3829		rg/gnccy	<u>More info</u>
DraIII	1	7080				cacnnn/gtg	<u>More info</u>
DrdI	3	1076	5143	7124		gacnnnn/nngtc	<u>More info</u>
DsaI	6	686	1062	3324	3424 4178 4514	c/ crygg	<u>More info</u>
EaeI	9	152	182	236	925 3298 4016 4273	y/ ggccr	<u>More info</u>
		4874	6316				
EagI	1	925				c/ ggccg	<u>More info</u>
Eam1104I	5	58	2482	2793	4918 6722	ctcttc	<u>More info</u>
Eam1105I	2	3754	5928			gacnnn/nngtc	<u>More info</u>
EaRI	5	58	2482	2793	4918 6722	ctcttc	<u>More info</u>
Ecl136II	1	892				gag/ ctc	<u>More info</u>
EclHKI	2	3754	5928			gacnnn/nngtc	<u>More info</u>
EclXI	1	925				c/ ggccg	<u>More info</u>
Eco105I	1	666				tac/gta	<u>More info</u>
Eco130I	11	686	1950	2226	3109 3324 3424	c/ cwwgg	<u>More info</u>
		3681	4060	4178	4514 4607		
Eco147I	2	3446	4606			agg/cct	<u>More info</u>
Eco24I	6	894	1017	1623	3526 3558 3806	grgcy/c	<u>More info</u>
Eco255I	1	6408				agt/act	<u>More info</u>
Eco31I	3	3380	4031	6000		ggtctc	<u>More info</u>
Eco32I	1	952				gat/ atc	<u>More info</u>
Eco52I	1	925				c/ ggccg	<u>More info</u>

168/173

FIG. 14-56

Eco57I	7	1210 2446 2488 3271 3314 5567 6615	ctgaag	<u>More info</u>
Eco64I	8	791 2264 3065 3434 3602 4779 5876 7036	g/ gyrcc	<u>More info</u>
Eco72I	1	2705	cac/gtg	<u>More info</u>
Eco81I	3	1034 1046 3256	cc/ tnagg	<u>More info</u>
Eco88I	3	3638 3934 4629	c/ ycgrg	<u>More info</u>
EcoICRI	1	892	gag/ ctc	<u>More info</u>
EcoNI	3	1259 1338 1684	cctnn/nnnagg	<u>More info</u>
EcoO109I	3	3291 3802 3829	rg/gnccy	<u>More info</u>
EcoRI	3	912 1990 2994	g/ aattc	<u>More info</u>
EcoRV	1	952	gat/ atc	<u>More info</u>
EcoT14I	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	<u>More info</u>
EcoT22I	3	3961 4356 4429	atgca/t	<u>More info</u>
ErhI	11	686 1950 2226 3109 3324 3424 3681 4060 4178 4514 4607	c/ cwwgg	<u>More info</u>
Esp1396I	6	1445 1482 1775 1796 2644 4191	ccannnn/ntgg	<u>More info</u>
Esp3I	3	2023 2773 4001	cgtctc	<u>More info</u>
FauNDI	1	560	ca/ tatg	<u>More info</u>
FbaI	1	969	t/ gatca	<u>More info</u>
FriOI	6	894 1017 1623 3526 3558 3806	grgcy/c	<u>More info</u>
FspI	2	21 6150	tgc/gca	<u>More info</u>

169/173

FIG. 14-57

170/173

GsuI	9	1015 1279 1772 2781 2842 3022	ctggag	<u>More info</u>
HaeII	5	3701 3863 6018		<u>More info</u>
HinII	6	2519 4913 5283 6922 6930	rgcgc/y	<u>More info</u>
HincII	3	448 501 584 770 4151 6465	gr/cgyc	<u>More info</u>
HindII	3	311 446 842	gty/rac	<u>More info</u>
HindIII	3	311 446 842	gty/rac	<u>More info</u>
Hsp92I	6	918 1394 2183	a/ agctt	<u>More info</u>
KpnI	3	448 501 584 770 4151 6465	gr/cgyc	<u>More info</u>
Ksp22I	1	2268 3438 3606	ggtac/c	<u>More info</u>
Ksp632I	5	969	t/ gatca	<u>More info</u>
LspI	3	58 2482 2793 4918 6722	ctcttc	<u>More info</u>
MfeI	1	1603 1988 2423	tt/cgaa	<u>More info</u>
MflI	12	1091	c/ aattg	<u>More info</u>
		932 2400 2634 3409 3596 3634	r/ gatcy	<u>More info</u>
		5676 5687 5773 5785 6553 6570		
MluNI	4	184 238 3300 4018	tgg/cca	<u>More info</u>
Mph1103I	3	3961 4356 4429	atgca/t	<u>More info</u>
MroNI	1	6972	g/ ccggc	<u>More info</u>
MscI	4	184 238 3300 4018	tgg/cca	<u>More info</u>
MslI	10	691 2094 2703 3323 3489 3651	caynn/nnrtg	<u>More info</u>
		3698 6180 6339 6698		
Msp17I	6	448 501 584 770 4151 6465	gr/cgyc	<u>More info</u>
MspA1I	7	71 2341 2731 4859 5377 5622 6563	cmg/ckg	<u>More info</u>

FIG. 14-58

171/173

MunI	1	1091		c/ aattg	More info
MvaI269I	1	1886		gaatgc	More info
NaeI	1	6974		gcc/ggc	More info
NcoI	5	686 3324 3424 4178 4514		c/ catgg	More info
NdeI	1	560		ca/ tatg	More info
NgoAIV	1	6972		g/ ccggc	More info
NgomI	1	6972		g/ ccggc	More info
NotI	1	925		gc/ggccgc	More info
NsiI	3	3961 4356 4429		atgca/t	More info
NspBII	7	71 2341 2731 4859 5377 5622 6563		cmg/ckg	More info
NspI	5	2930 3959 4354 4427 5039		rcatg/y	More info
NspV	3	1603 1988 2423		tt/cgaa	More info
PaeI	4	2930 3959 4354 4427		gcatg/c	More info
Paer7I	1	4629		c/ tcgag	More info
PflMI	6	1445 1482 1775 1796 2644 4191		ccannnn/ntgg	More info
PinAI	1	4188		a/ ccggt	More info
Ple19I	2	42 6298		cgat/cg	More info
PmaCI	1	2705		cac/gtg	More info
Pme55I	2	3446 4606		agg/cct	More info
PmlI	1	2705		cac/gtg	More info
Ppu10I	3	3957 4352 4425		a/ tgcat	More info
PshBI	4	334 4806 4865 6100		at/ taat	More info
Psp124BI	1	894		gagct/c	More info
Psp1406I	3	6154 6527 7291		aa/cggt	More info

FIG. 14-59

172/173

PspAI	1	3638	c/ ccggg	<u>More info</u>
PspALI	1	3640	ccc/ggg	<u>More info</u>
PspOMI	1	3802	g/ ggccc	<u>More info</u>
PstI	2	948 2148	ctgca/g	<u>More info</u>
PvuI	2	42 6298	cgat/cg	<u>More info</u>
PvuII	3	71 2341 4859	cag/ctg	<u>More info</u>
RcaI	3	1891 5755 6763	t/ catga	<u>More info</u>
SacI	1	894	gagct/c	<u>More info</u>
SapI	2	2483 4918	gctcttc	<u>More info</u>
ScaI	1	6408	agt/act	<u>More info</u>
SexAI	1	4373	a/ ccwggg	<u>More info</u>
SfcI	8	944 2144 3824 4662 5300 5491 6169 6854	c/ tryag	<u>More info</u>
SfiI	1	4560	ggccnnnn/nggcc	<u>More info</u>
Sfr274I	1	4629	c/ tcgag	<u>More info</u>
SfuI	3	1603 1988 2423	tt/cgaa	<u>More info</u>
SmaI	1	3640	ccc/ggg	<u>More info</u>
SnaBI	1	666	tac/gta	<u>More info</u>
SpeI	1	326	a/ ctagt	<u>More info</u>
SphI	4	2930 3959 4354 4427	gcatg/c	<u>More info</u>
SseBI	2	3446 4606	agg/cct	<u>More info</u>
SspBI	2	270 3471	t/ gtaca	<u>More info</u>
SspI	5	179 226 3768 6732 7285	aat/att	<u>More info</u>
SstI	1	894	gagct/c	<u>More info</u>

FIG. 14-60

StuI	2	3446	4606					agg/cct	<u>More info</u>
StyI	11	686	1950	2226	3109	3324	3424	c/ cwwgg	<u>More info</u>
		3681	4060	4178	4514	4607			
Van91I	6	1445	1482	1775	1796	2644	4191	ccannnn/ntgg	<u>More info</u>
VneI	2	5349	6595					g/ tgcac	<u>More info</u>
VspI	4	334	4806	4865	6100			at/ taat	<u>More info</u>
XcmI	2	1948	2897					ccannnn/nnntgg	<u>More info</u>
XhoI	1	4629						c/ tcgag	<u>More info</u>
XhoII	12	932	2400	2634	3409	3596	3634	r/ gatcy	<u>More info</u>
		5676	5687	5773	5785	6553	6570		
XmaI	1	3638						c/ ccggg	<u>More info</u>
XmaIII	1	925						c/ ggccg	<u>More info</u>
XmnI	4	1107	2481	3506	6527			gaann/nnttc	<u>More info</u>
Zsp2I	3	3961	4356	4429				atgca/t	<u>More info</u>

173/173

The following endonucleases were selected but don't cut this sequence:

AccI, AccIII, AfeI, AflII, Aor51HI, AscI, AspI, AtsI, BbeI, BfrI, BsaBI, Bse8I, BseAI, BsePI, Bsh1365I, BsiMI, BsiWI, Bsp13I, Bsp68I, BspEI, BspTI, BsrBRI, BssHII, Bst1107I, Bst98I, BstEII, BstPI, Cfr42I, CpoI, CspI, Eco47III, Eco91I, EcoO65I, EheI, FseI, HpaI, Kasi, Kpn2I, KspI, Mami, MluI, MroI, MspCI, Nari, NheI, NruI, PacI, Pfl23II, PmeI, PpuMI, PshAI, Psp5II, PspEI, PspLI, PstNHI, RsrII, SacII, Sali, SbfI, Sfr303I, Sgfi, SgrAI, SmaI, SmlI, Sphi, SrfI, Sse8387I, SstII, SunI, SwaI, Tth111I, Vha464I, XbaI

FIG. 14-61



1/25

## SEQUENCE LISTING

<110> Sloan-Kettering Institute for Cancer Research  
Richon, Victoria  
Zhou, Xianbo  
Rifkind, Richard A.  
Marks, Paul A.

<120> HDAC9 Polypeptides and Polynucleotides  
and Uses Thereof

<130> 3254.1000005

<150> 60/298,173

<151> 2001-06-14

<150> 60/311,686

<151> 2001-08-10

<150> 60/316,995

<151> 2001-09-04

<160> 22

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 3186

<212> DNA

<213> Homo sapiens

<400> 1

ggggaagaga	ggcacagaca	cagataggag	aagggcaccg	gctggagcca	cttgcaggac	60
tgagggtttt	tgcaacaaaa	ccctagcagc	ctgaagaact	ctaagccaga	tggggtggct	120
ggacgagagc	agctcttggc	tcagcaaaga	atgcacagta	tgatcagctc	agtggatgtg	180
aagtcagaag	ttctgtggg	cctggagccc	atctcacctt	tagacctaa	gacagacctc	240
aggatgatga	tgcccggtg	ggacctgtt	gtccgtgaga	agcaattgca	gcaggaatta	300
cttcttatcc	agcagcagca	acaaatccag	aagcagcttc	tgatagcaga	gtttcagaaa	360
cagcatgaga	acttgacacg	gcagcaccag	gctcagcttc	aggagcatat	caaggaactt	420
ctagccataa	aacagcaaca	agaactccta	gaaaaggagc	agaaactgga	gcagcagagg	480
caagaacagg	aagtagagag	gcacgcagca	gaacagcagc	ttcctcctct	cagaggcaaa	540
gatagaggac	gagaaagggc	agtggcaagt	acagaagtaa	agcagaagct	tcaagagttc	600
ctactgagta	aatcagcaac	gaaagacact	ccaactaatg	gaaaaaatca	ttccgtgagc	660
cgccatccca	agctctggta	cacggctgcc	caccacacat	cattggatca	aagctctcca	720
ccccttagtg	gaacatctcc	atctacaag	tacacattac	caggagcaca	agatgcaaag	780
gatgatttcc	cccttcgaaa	aactgcctct	gagcccaact	tgaagggtcg	gtccagggtta	840
aaacagaaa	tgccagagag	gagaagcagc	cccttactca	ggcggaagga	tggaaatggt	900
gtcacttcat	tcaagaagcg	aatgtttgag	gtgacagaat	cctcagtcag	tagcagttct	960
ccaggctctg	gtcccagttc	accaaacaat	gggccaaactg	gaagtgttac	tgaaaatgag	1020
acttcggttt	tgccccctac	ccctcatgcc	gagcaaatgg	tttcacagca	acgcattcta	1080
attcatgaag	attccatgaa	cctgctaagt	ctttatacct	ctccttcttt	gccaacatt	1140
accttggggc	ttcccgagc	gccatcccag	ctcaatgctt	cgaattcact	caaagaaaag	1200
cagaagtgtg	agacgcagac	gcttaggcaa	ggtgttcctc	tgccctgggca	gtatggaggc	1260
agcatcccg	cactctccag	ccaccctcat	gttactttag	agggaaaagcc	acccaacagc	1320
agccaccagg	ctctcctgca	gcatttatta	ttgaaagaac	aaatgcgaca	gcaaaagctt	1380
cttgtagctg	gtggagttcc	cttacatcct	cagctctcct	tggcaacaaa	agagagaatt	1440
tcacctggca	ttagagggtac	ccacaaattg	ccccgtcaca	gacccctgaa	ccgaacccag	1500
tctgcacctt	tgccctcagag	caggttggtc	cagctggtca	ttcaacagca	acaccagcaa	1560
ttcttggaga	agcagaagca	ataccagcag	cagatccaca	tgaacaaact	gctttcgaaa	1620
tctattgaac	aactgaagca	accaggcagt	caccttgagg	aagcagagga	agagcttcag	1680
ggggaccagg	cgatgcagga	agacagagcg	ccctctagt	gcaacagcac	taggagcgac	1740

2/25

```

agcagtgcct gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac agcctttcct ggaacccacg cacacacgtg cgctctctgt gcgccaagct 1920
ccgtgggtcg cggttggcat ggatggatta gagaacacc gtctcgtctc caggactcac 1980
tcttccctcg ctgcctctgt tttacctcac ccagcaatgg accgccccct ccagcctggc 2040
tctgcaactg gaattgccta tgacccttg atgctgaaac accagtgcgt ttgtggcaat 2100
tccaccaccc accctgagca tgctggacga atacagagta tctggtcacg actgcaagaa 2160
actgggctgc taaataaatg tgagcgaatt caaggtcgaa aagccagcct ggaggaaata 2220
cagcttggtc attctgaaca tcaactactg ttgtatggca ccaacccccct ggacggacag 2280
aagctggacc ccaggatact cctagtgatg gactctcaaa agtttttttc ctcattacct 2340
tgtgtgggac ttgggtggga cagtgcacac atttgggaatg agctacactc gtccggtgct 2400
gcacgcatgg ctgttggctg tgtcatcgag ctggcttcca aagtggcctc aggagagctg 2460
aagaatgggt ttgctgttgt gaggccccct ggccatcacg ctgaagaatc cacagccatg 2520
gggttctgct tttttaattc agttgcaatt accgccaat acttgagaga ccaactaaat 2580
ataagcaaga tattgattgt agatctggat gttcaccatg gaaacggtac ccagcaggcc 2640
ttttatgctg accccagcat cctgtacatt tcaactccatc gctatgatga agggaaacttt 2700
ttccctggca gtggagcccc aaatgaggtt ggaacaggcc ttggagaagg gtacaatata 2760
aatattgcct ggacaggtgg ccttgatcct cccatgggag atgttgagta ccttgaagca 2820
ttcaggacca tcgtgaagcc tgtggccaaa gagtttgatc cagacatggt cttagtatct 2880
gctggatttg atgcattgga aggccacacc cctcctctag gaggtacaa agtgacggca 2940
aaatgttttg gtcatttgac gaagcaattg atgacattgg ctgatggacg tgtggtgttg 3000
gctctagaag gaggacatga tctcacagcc atctgtgatg catcagaagc ctgtgtaaat 3060
gcccttctag gaaatgagct ggagccactt gcagaagata ttctccacca aagcccgaat 3120
atgaatgctg ttatttcttt acagaagatc attgaaattc aaagtatgtc tttaaagttc 3180
tcttaa

```

&lt;210&gt; 2

&lt;211&gt; 1011

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130 135 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145 150 155 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165 170 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180 185 190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195 200 205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
210 215 220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225 230 235 240

```

3/25

Pro	Leu	Leu	Arg	Arg	Lys	Asp	Gly	Asn	Val	Val	Thr	Ser	Phe	Lys	Lys
				245					250					255	2
Arg	Met	Phe	Glu	Val	Thr	Glu	Ser	Ser	Val	Ser	Ser	Ser	Ser	Pro	Gly
			260					265					270		
Ser	Gly	Pro	Ser	Ser	Pro	Asn	Asn	Gly	Pro	Thr	Gly	Ser	Val	Thr	Glu
		275					280					285			
Asn	Glu	Thr	Ser	Val	Leu	Pro	Pro	Thr	Pro	His	Ala	Glu	Gln	Met	Val
	290					295				300					
Ser	Gln	Gln	Arg	Ile	Leu	Ile	His	Glu	Asp	Ser	Met	Asn	Leu	Leu	Ser
305				310						315					320
Leu	Tyr	Thr	Ser	Pro	Ser	Leu	Pro	Asn	Ile	Thr	Leu	Gly	Leu	Pro	Ala
				325					330					335	
Val	Pro	Ser	Gln	Leu	Asn	Ala	Ser	Asn	Ser	Leu	Lys	Glu	Lys	Gln	Lys
			340					345					350		
Cys	Glu	Thr	Gln	Thr	Leu	Arg	Gln	Gly	Val	Pro	Leu	Pro	Gly	Gln	Tyr
		355					360					365			
Gly	Gly	Ser	Ile	Pro	Ala	Ser	Ser	Ser	His	Pro	His	Val	Thr	Leu	Glu
370						375					380				
Gly	Lys	Pro	Pro	Asn	Ser	Ser	His	Gln	Ala	Leu	Leu	Gln	His	Leu	Leu
385				390						395					400
Leu	Lys	Glu	Gln	Met	Arg	Gln	Gln	Lys	Leu	Leu	Val	Ala	Gly	Gly	Val
				405					410					415	
Pro	Leu	His	Pro	Gln	Ser	Pro	Leu	Ala	Thr	Lys	Glu	Arg	Ile	Ser	Pro
			420					425					430		
Gly	Ile	Arg	Gly	Thr	His	Lys	Leu	Pro	Arg	His	Arg	Pro	Leu	Asn	Arg
		435					440					445			
Thr	Gln	Ser	Ala	Pro	Leu	Pro	Gln	Ser	Thr	Leu	Ala	Gln	Leu	Val	Ile
450						455					460				
Gln	Gln	Gln	His	Gln	Gln	Phe	Leu	Glu	Lys	Gln	Lys	Gln	Tyr	Gln	Gln
465				470						475					480
Gln	Ile	His	Met	Asn	Lys	Leu	Leu	Ser	Lys	Ser	Ile	Glu	Gln	Leu	Lys
				485					490					495	
Gln	Pro	Gly	Ser	His	Leu	Glu	Glu	Ala	Glu	Glu	Glu	Leu	Gln	Gly	Asp
			500					505					510		
Gln	Ala	Met	Gln	Glu	Asp	Arg	Ala	Pro	Ser	Ser	Gly	Asn	Ser	Thr	Arg
		515					520					525			
Ser	Asp	Ser	Ser	Ala	Cys	Val	Asp	Asp	Thr	Leu	Gly	Gln	Val	Gly	Ala
		530				535					540				
Val	Lys	Val	Lys	Glu	Glu	Pro	Val	Asp	Ser	Asp	Glu	Asp	Ala	Gln	Ile
545				550						555					560
Gln	Glu	Met	Glu	Ser	Gly	Glu	Gln	Ala	Ala	Phe	Met	Gln	Gln	Pro	Phe
				565					570					575	
Leu	Glu	Pro	Thr	His	Thr	Arg	Ala	Leu	Ser	Val	Arg	Gln	Ala	Pro	Leu
			580					585					590		
Ala	Ala	Val	Gly	Met	Asp	Gly	Leu	Glu	Lys	His	Arg	Leu	Val	Ser	Arg
		595					600					605			
Thr	His	Ser	Ser	Pro	Ala	Ala	Ser	Val	Leu	Pro	His	Pro	Ala	Met	Asp
					615						620				
Arg	Pro	Leu	Gln	Pro	Gly	Ser	Ala	Thr	Gly	Ile	Ala	Tyr	Asp	Pro	Leu
625				630						635					640
Met	Leu	Lys	His	Gln	Cys	Val	Cys	Gly	Asn	Ser	Thr	Thr	His	Pro	Glu
				645					650					655	
His	Ala	Gly	Arg	Ile	Gln	Ser	Ile	Trp	Ser	Arg	Leu	Gln	Glu	Thr	Gly
			660					665					670		
Leu	Leu	Asn	Lys	Cys	Glu	Arg	Ile	Gln	Gly	Arg	Lys	Ala	Ser	Leu	Glu
		675						680				685			
Glu	Ile	Gln	Leu	Val	His	Ser	Glu	His	His	Ser	Leu	Leu	Tyr	Gly	Thr
					695						700				
Asn	Pro	Leu	Asp	Gly	Gln	Lys	Leu	Asp	Pro	Arg	Ile	Leu	Leu	Gly	Asp
705					710					715					720
Asp	Ser	Gln	Lys	Phe	Ser	Ser	Leu	Pro	Cys	Gly	Gly	Leu	Gly	Val	
				725				730						735	

4/25

Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Ala Arg  
 740 745 750  
 Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly  
 755 760 765  
 Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala  
 770 775 780  
 Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile  
 785 790 795 800  
 Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile  
 805 810 815  
 Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr  
 820 825 830  
 Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly  
 835 840 845  
 Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Gly Thr Gly Leu  
 850 855 860  
 Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly Gly Leu Asp Pro  
 865 870 875 880  
 Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg Thr Ile Val Lys  
 885 890 895  
 Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu Val Ser Ala Gly  
 900 905 910  
 Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly Gly Tyr Lys Val  
 915 920 925  
 Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu Met Thr Leu Ala  
 930 935 940  
 Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His Asp Leu Thr Ala  
 945 950 955 960  
 Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu Leu Gly Asn Glu  
 965 970 975  
 Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser Pro Asn Met Asn  
 980 985 990  
 Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln Ser Met Ser Leu  
 995 1000 1005  
 Lys Phe Ser  
 1010

<210> 3  
 <211> 3499  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttctctgtggg cctggagccc atctcacctt tagacctaa gacagacctc 240  
 aggatgatga tgcccgtggg ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaattccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
 ctaccataaa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgagag gaacagcagc acagaagtaa agcagaagct tcaagagttc 600  
 gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaaag 780  
 gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgag gtccaggtta 840  
 aaacagaaag tggcagagag gagaagcagc cccttactca ggcggaagga tggaaatgtt 900  
 gtcaacttcac tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960  
 ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020  
 acttcggttt tgccccctac ccctcatgcc gagcaaatgg ttccacagca acgcattcta 1080

5/25

```

attcatgaag attccatgaa cctgctaagt ctttatacct ctccttcttt gcccaacatt 1140
accttggggc ttcccgcagt gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgcctgggca gtatggaggc 1260
agcatcccg gcatctccag ccaccctcat gttactttag agggaaagcc acccaacagc 1320
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380
ctttagactg gtggagttcc cttacatcct cagtctccct tggcaacaaa agagagaatt 1440
tcacctggca ttagaggtag ccacaaattg ccccgtcaca gacccttgaa ccgaaccag 1500
tctgcacctt tgcctcagag cacgttggtc cagctgggtc ttcaacagca acaccagcaa 1560
ttcttgagga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680
ggggaccagg cgatgcagga agacagagcg ccctctagt gcaacagcac taggagcgac 1740
agcagtgcct gtgtggatga cacactggga caagttgggg ctgtgaaggt caaggaggaa 1800
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860
tttatgcaac agcctttcct ggaacccacg cacacacgtg cgctctctgt gcgccaagct 1920
ccgctggctg cggttggcat ggatggatta gagaaacacc gtctcgtctc caggactcac 1980
tcttcccctg ctgctctgt tttacctcac ccagcaatgg accgccccct ccagcctggc 2040
tctgcaactg gaattgccta tgacccttg atgtgaaac accagtgcgt ttgtggcaat 2100
tccaccacc accctgagca tgctggacga atacagagta tctggtcacg actgcaagaa 2160
actgggctgc taaataaatg tgagcgaatt caaggtcgaa aagccagcct ggaggaaata 2220
cagcttggtt attctgaaca tcaactcactg ttgtatggca ccaacccccct ggacggacag 2280
aagctggacc ccaggatact cctaggtgat gactctcaaa agtttttttc ctcatctacc 2340
tgtgtgggac ttgggttgga cagtgcaccc atttggaatg agctacactc gtccgggtgct 2400
gcacgcattg ctgttggtg tgctatcgag ctggcttcca aagtggcctc aggagagctg 2460
aagaatgggt ttgctgtgt gaggccccct ggccatcacg ctgaagaatc cacagccatg 2520
gggttctgct tttttaattc agttgcaatt accgccaat acttgagaga ccaactaaat 2580
ataagcaaga tattgattgt agatctggat gttcaccatg gaaacggtac ccagcaggcc 2640
ttttatgctg accccagcat cctgtacatt tcaactccatc gctatgatga agggaaactt 2700
ttccctggca gtggagcccc aaatgaggtt cggtttattt ctttagagcc ccacttttat 2760
ttgatctttt caggttaattg cattgcatag ttaccctaa ttttcttgct ctttgctggt 2820
gttttaaat acacgagatt actgaattgt cccatgggac caagaaccag tgcagaacaa 2880
gtgcataacc cagagcactg tttgtcaggg aaggttgggc tgatttgatg tgttggttga 2940
tgtttatttc aagagctccc atgtgcttgt tttcctctct tcttgctttc ttccatttgc 3000
tctcttctct gccaccgtg gtgtgtctt ctcttccag gttggaacag gccttgagga 3060
aggtacaat ataaatattg cctggacagg tggccttgat cctcccatgg gagatgttga 3120
gtaccttgaa gcattcagga ccactcgtgaa gcctgtggcc aaagagtttg atccagacat 3180
ggtcttagta tctgctggat ttgatgcatt ggaaggccac acccctcctc taggagggtg 3240
caaagtgcag gcaaaatgtt ttggtcattt gacgaagcaa ttgatgacat tggctgatgg 3300
acgtgtggtg ttggtcttag aaggaggaca tgatctcaca gccatctgtg atgcatcaga 3360
agcctgtgta aatgcccttc taggaaatga gctggagcca cttgcagaag atattctcca 3420
ccaaagcccg aatatgaatg ctgttatttc tttacagaag atcattgaaa ttcaaagtat 3480
gtctttaaag ttctcttaa 3499

```

&lt;210&gt; 4

&lt;211&gt; 879

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 4

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110

```

6/25

Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg  
 115 120 125  
 Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys  
 130 135 140  
 Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr  
 145 150 155 160  
 Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp  
 165 170 175  
 Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu  
 180 185 190  
 Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp  
 195 200 205  
 Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu  
 210 215 220  
 Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser  
 225 230 235 240  
 Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys  
 245 250 255  
 Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly  
 260 265 270  
 Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu  
 275 280 285  
 Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val  
 290 295 300  
 Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser  
 305 310 315 320  
 Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala  
 325 330 335  
 Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys  
 340 345 350  
 Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr  
 355 360 365  
 Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu  
 370 375 380  
 Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu  
 385 390 395 400  
 Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val  
 405 410 415  
 Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro  
 420 425 430  
 Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg  
 435 440 445  
 Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile  
 450 455 460  
 Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln  
 465 470 475 480  
 Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys  
 485 490 495  
 Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp  
 500 505 510  
 Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg  
 515 520 525  
 Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala  
 530 535 540  
 Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile  
 545 550 555 560  
 Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Pro Phe  
 565 570 575  
 Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg Gln Ala Pro Leu  
 580 585 590  
 Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg Leu Val Ser Arg  
 595 600 605

7/25

Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His Pro Ala Met Asp  
 610 615 620  
 Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala Tyr Asp Pro Leu  
 625 630 635 640  
 Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr Thr His Pro Glu  
 645 650 655  
 His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu Gln Glu Thr Gly  
 660 665 670  
 Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys Ala Ser Leu Glu  
 675 680 685  
 Glu Ile Gln Leu Val His Ser Glu His His Ser Leu Leu Tyr Gly Thr  
 690 695 700  
 Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile Leu Leu Gly Asp  
 705 710 715 720  
 Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly Gly Leu Gly Val  
 725 730 735  
 Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Arg  
 740 745 750  
 Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly  
 755 760 765  
 Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala  
 770 775 780  
 Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile  
 785 790 795 800  
 Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile  
 805 810 815  
 Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr  
 820 825 830  
 Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly  
 835 840 845  
 Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Arg Phe Ile Ser  
 850 855 860  
 Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn Cys Ile Ala  
 865 870 875

&lt;210&gt; 5

&lt;211&gt; 3054

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctaggcagc ctgaagaact ctaagccaga tggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttctgtgtgg cctggagccc atctcacctt tagacctaat gacagacctc 240  
 aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgaga gaacagcagc ttctctctct cagaggcaaa 540  
 gatagaggac gagaaagggc agtggaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggtt cacggctgcc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaaag 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840  
 cccagttcac caacaatgg gccactgga agtggtactg aaaatgagac ttcggttttg 900  
 cccctaccc ctcattgccg gcaaatggtt tcacagcaac gcattctaatt tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020  
 cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080  
 acgcagacgc ttaggcaagg tggtcctctg cctgggcagt atggaggcag catcccgca 1140  
 tcttcagcc accctcatgt tacttttagg ggaaagccac ccaacagcag ccaccaggct 1200

```

ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtagcc acaaatggcc ccgtcacaga cccctgaacc gaaccagtc tgcacctttg 1380
cctcagagca cgttggtcga gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtc ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgccttg 1620
gtggatgaca cactgggaca agttggggct gtgaaggcca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggtctgcttt tatgcaacag 1740
cctttcctgg aaccacagca cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctctgtctcca ggactcactc tccccctgct 1860
gcctctgttt tacctcaccg agcaatggac cgccccctcc agcctggctc tgcaactgga 1920
attgcctatg accccttgat gctgaaacac cagtgcgttt gtggcaattc caccaccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaatata gcttggtcat 2100
tctgaacatc actcactgtt gtatggcacc aacccccctg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaaag tttttttcct cattaccttg tgggtggactt 2220
ggggtggaca gtgacacatc ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgtgtgta ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtagcc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagcccaaa atgaggttgg aacaggcctt ggagaagggt acaatataaa tattgcctgg 2640
acaggtggcc ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc 2700
gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc tggatttgat 2760
gcattggaag gccacacccc tcctctagga gggtacaaa tgacggcaaa atgttttggt 2820
catttgacga agcaattgat gacattggct gatggacgtg tgggtgttggc tctagaagga 2880
ggacatgatc tcacagccat ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga 2940
aatgagctgg agccacttgc agaagatatt ctccacaaa gcccgaaat gatgtgctt 3000
atttctttac agaagatcat tgaaattcaa agtatgtctt taaagtcttc ttaa 3054

```

<210> 6  
 <211> 967  
 <212> PRT  
 <213> Homo sapiens

<400> 6  
 Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val  
 1 5 10 15  
 Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met  
 20 25 30  
 Met Met Pro Val Val Asp Pro Val Arg Glu Lys Gln Leu Gln Gln  
 35 40 45  
 Glu Leu Leu Leu Ile Gln Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu  
 50 55 60  
 Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln  
 65 70 75 80  
 Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln  
 85 90 95  
 Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu  
 100 105 110  
 Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg  
 115 120 125  
 Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys  
 130 135 140  
 Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr  
 145 150 155 160  
 Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp  
 165 170 175  
 Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu  
 180 185 190



9/25

Ser	Gly	Thr	Ser	Pro	Ser	Tyr	Lys	Tyr	Thr	Leu	Pro	Gly	Ala	Gln	Asp
		195					200					205			
Ala	Lys	Asp	Asp	Phe	Pro	Leu	Arg	Lys	Thr	Glu	Ser	Ser	Val	Ser	Ser
	210					215					220				
Ser	Ser	Pro	Gly	Ser	Gly	Pro	Ser	Ser	Pro	Asn	Asn	Gly	Pro	Thr	Gly
225					230					235					240
Ser	Val	Thr	Glu	Asn	Glu	Thr	Ser	Val	Leu	Pro	Pro	Thr	Pro	His	Ala
				245					250					255	
Glu	Gln	Met	Val	Ser	Gln	Gln	Arg	Ile	Leu	Ile	His	Glu	Asp	Ser	Met
			260					265					270		
Asn	Leu	Leu	Ser	Leu	Tyr	Thr	Ser	Pro	Ser	Leu	Pro	Asn	Ile	Thr	Leu
		275					280					285			
Gly	Leu	Pro	Ala	Val	Pro	Ser	Gln	Leu	Asn	Ala	Ser	Asn	Ser	Leu	Lys
	290					295					300				
Glu	Lys	Gln	Lys	Cys	Glu	Thr	Gln	Thr	Leu	Arg	Gln	Gly	Val	Pro	Leu
305					310					315					320
Pro	Gly	Gln	Tyr	Gly	Gly	Ser	Ile	Pro	Ala	Ser	Ser	Ser	His	Pro	His
				325					330					335	
Val	Thr	Leu	Glu	Gly	Lys	Pro	Pro	Asn	Ser	Ser	His	Gln	Ala	Leu	Leu
			340					345					350		
Gln	His	Leu	Leu	Leu	Lys	Glu	Gln	Met	Arg	Gln	Gln	Lys	Leu	Leu	Val
		355					360					365			
Ala	Gly	Gly	Val	Pro	Leu	His	Pro	Gln	Ser	Pro	Leu	Ala	Thr	Lys	Glu
	370					375					380				
Arg	Ile	Ser	Pro	Gly	Ile	Arg	Gly	Thr	His	Lys	Leu	Pro	Arg	His	Arg
385					390					395					400
Pro	Leu	Asn	Arg	Thr	Gln	Ser	Ala	Pro	Leu	Pro	Gln	Ser	Thr	Leu	Ala
				405					410					415	
Gln	Leu	Val	Ile	Gln	Gln	His	Gln	Gln	Phe	Leu	Glu	Lys	Gln	Lys	
			420					425					430		
Gln	Tyr	Gln	Gln	Gln	Ile	His	Met	Asn	Lys	Leu	Leu	Ser	Lys	Ser	Ile
		435					440					445			
Glu	Gln	Leu	Lys	Gln	Pro	Gly	Ser	His	Leu	Glu	Glu	Ala	Glu	Glu	Glu
	450					455					460				
Leu	Gln	Gly	Asp	Gln	Ala	Met	Gln	Glu	Asp	Arg	Ala	Pro	Ser	Ser	Gly
465					470					475					480
Asn	Ser	Thr	Arg	Ser	Asp	Ser	Ser	Ala	Cys	Val	Asp	Asp	Thr	Leu	Gly
				485					490					495	
Gln	Val	Gly	Ala	Val	Lys	Val	Lys	Glu	Glu	Pro	Val	Asp	Ser	Asp	Glu
			500					505					510		
Asp	Ala	Gln	Ile	Gln	Glu	Met	Glu	Ser	Gly	Glu	Gln	Ala	Ala	Phe	Met
		515					520					525			
Gln	Gln	Pro	Phe	Leu	Glu	Pro	Thr	His	Thr	Arg	Ala	Leu	Ser	Val	Arg
		530				535					540				
Gln	Ala	Pro	Leu	Ala	Ala	Val	Gly	Met	Asp	Gly	Leu	Glu	Lys	His	Arg
545					550					555					560
Leu	Val	Ser	Arg	Thr	His	Ser	Ser	Pro	Ala	Ala	Ser	Val	Leu	Pro	His
				565					570					575	
Pro	Ala	Met	Asp	Arg	Pro	Leu	Gln	Pro	Gly	Ser	Ala	Thr	Gly	Ile	Ala
			580					585					590		
Tyr	Asp	Pro	Leu	Met	Leu	Lys	His	Gln	Cys	Val	Cys	Gly	Asn	Ser	Thr
		595					600					605			
Thr	His	Pro	Glu	His	Ala	Gly	Arg	Ile	Gln	Ser	Ile	Trp	Ser	Arg	Leu
		610				615					620				
Gln	Glu	Thr	Gly	Leu	Leu	Asn	Lys	Cys	Glu	Arg	Ile	Gln	Gly	Arg	Lys
625					630					635					640
Ala	Ser	Leu	Glu	Glu	Ile	Gln	Leu	Val	His	Ser	Glu	His	His	Ser	Leu
				645					650					655	
Leu	Tyr	Gly	Thr	Asn	Pro	Leu	Asp	Gly	Gln	Lys	Leu	Asp	Pro	Arg	Ile
			660					665					670		
Leu	Leu	Gly	Asp	Asp	Ser	Gln	Lys	Phe	Phe	Ser	Ser	Leu	Pro	Cys	Gly
		675					680						685		

10/25

Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser  
 690 695 700  
 Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys  
 705 710 715 720  
 Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro  
 725 730 735  
 Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn  
 740 745 750  
 Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser  
 755 760 765  
 Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln  
 770 775 780  
 Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg  
 785 790 795 800  
 Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val  
 805 810 815  
 Gly Thr Gly Leu Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly  
 820 825 830  
 Gly Leu Asp Pro Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg  
 835 840 845  
 Thr Ile Val Lys Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu  
 850 855 860  
 Val Ser Ala Gly Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly  
 865 870 875 880  
 Gly Tyr Lys Val Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu  
 885 890 895  
 Met Thr Leu Ala Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His  
 900 905 910  
 Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu  
 915 920 925  
 Leu Gly Asn Glu Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser  
 930 935 940  
 Pro Asn Met Asn Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln  
 945 950 955 960  
 Ser Met Ser Leu Lys Phe Ser  
 965

&lt;210&gt; 7

&lt;211&gt; 3367

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 7

ggggaagaga ggacacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttctgtggg cctggagccc atctcacctt tagacctaaag gacagacctc 240  
 aggatgatga tgcccgtggg ggaccctggt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgagca gaacagcagc ttctctctct cagaggcaaa 540  
 gatagaggac gagaaagggc agtggaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggta cacgggtgcc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atctacaag tacacattac caggagcaca agatgcaaa 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggcctctgg 840  
 cccagttcac caaacaatgg gccaaactgga agtggtactg aaaatgagac ttccgttttg 900  
 cccctacccc catcgccga gcaaatgggt tcacagcaac gcattctaatt tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020  
 cccgcagtgc catccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080

11/25

```

acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catccccgga 1140
tcttcacagcc accctcatgt tacttttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcctct tgtagctggt 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc ccgtcacaga cccctgaacc gaaccagtc tgcacctttg 1380
cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtc ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcctgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggta aggaggaacc agtggacagt 1680
gatgaagatg ctcatatcca ggaaatggaa tctggggagc aggtctgctt tatgcaacag 1740
cctttccttg aaccacagca cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctctctcca ggactcactc ttcccctgct 1860
gcctctggtt tacctcaccg agcaatggac cgcccctcc agcctggctc tgcaactgga 1920
attgcctatg accccttgat gctgaaacac cagtgcgttt gtggcaattc caccaccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttggtcat 2100
tctgaacatc actcactggt gtatggcacc aaccctctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag ttttttctc cattaccttg tgggtgacct 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggtctgt tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgatg ggccccctgg ccatcacgct gaagaatcca cagccatggg gttctgctt 2400
ttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtaccc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagcccca atgaggttcg gtttatttct ttagagcccc acttttattt gtatctttca 2640
ggttaattgca ttgcatgatt acccctaatt ttcttgcct ttgctggtgt tttaaattac 2700
acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt gcataaccga 2760
gagcactggt tctcagggaa gtttggctg atttgatgt ttgtttgatg tttatttcaa 2820
gagctcccat gtgcttgttt tctctcttc ttgctttctt ccatttgctc tcttctctgc 2880
ccaccgtggt gtgtctttct cttcccagggt tggaaacaggc cttggagaag ggtacaatat 2940
aaatattgcc tggacagggt gccttgatcc tcccatggga gatgttgagt accttgaagc 3000
attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg tcttagtatc 3060
tgctggattt gatgcattgg aaggccacac ccctcctcta ggagggtaca aagtgcaggc 3120
aaaatgtttt ggtcatttga cgaagcaatt gatgacattg gctgatggac gtgtggtgtt 3180
ggctctagaa ggaggacatg atctcacagc catctgtgat gcatcagaag cctgtgtaaa 3240
tgccttctta ggaatgagc tggagccact tgcagaagat attctccacc aaagcccga 3300
tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt ctttaaagtt 3360
ctcttaa 3367

```

&lt;210&gt; 8

&lt;211&gt; 835

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 8

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1      5      10      15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20      25      30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35      40      45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50      55      60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65      70      75      80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85      90      95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100     105     110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115     120     125

```

12/25

Gly	Lys	Asp	Arg	Gly	Arg	Glu	Arg	Ala	Val	Ala	Ser	Thr	Glu	Val	Lys
	130					135					140				
Gln	Lys	Leu	Gln	Glu	Phe	Leu	Leu	Ser	Lys	Ser	Ala	Thr	Lys	Asp	Thr
145					150					155					160
Pro	Thr	Asn	Gly	Lys	Asn	His	Ser	Val	Ser	Arg	His	Pro	Lys	Leu	Trp
				165					170						175
Tyr	Thr	Ala	Ala	His	His	Thr	Ser	Leu	Asp	Gln	Ser	Ser	Pro	Pro	Leu
			180					185					190		
Ser	Gly	Thr	Ser	Pro	Ser	Tyr	Lys	Tyr	Thr	Leu	Pro	Gly	Ala	Gln	Asp
		195					200					205			
Ala	Lys	Asp	Asp	Phe	Pro	Leu	Arg	Lys	Thr	Glu	Ser	Ser	Val	Ser	Ser
	210					215					220				
Ser	Ser	Pro	Gly	Ser	Gly	Pro	Ser	Ser	Pro	Asn	Asn	Gly	Pro	Thr	Gly
225					230					235					240
Ser	Val	Thr	Glu	Asn	Glu	Thr	Ser	Val	Leu	Pro	Pro	Thr	Pro	His	Ala
				245					250					255	
Glu	Gln	Met	Val	Ser	Gln	Gln	Arg	Ile	Leu	Ile	His	Glu	Asp	Ser	Met
		260						265					270		
Asn	Leu	Leu	Ser	Leu	Tyr	Thr	Ser	Pro	Ser	Leu	Pro	Asn	Ile	Thr	Leu
		275					280					285			
Gly	Leu	Pro	Ala	Val	Pro	Ser	Gln	Leu	Asn	Ala	Ser	Asn	Ser	Leu	Lys
	290					295					300				
Glu	Lys	Gln	Lys	Cys	Glu	Thr	Gln	Thr	Leu	Arg	Gln	Gly	Val	Pro	Leu
305					310					315					320
Pro	Gly	Gln	Tyr	Gly	Gly	Ser	Ile	Pro	Ala	Ser	Ser	Ser	His	Pro	His
				325					330					335	
Val	Thr	Leu	Glu	Gly	Lys	Pro	Pro	Asn	Ser	Ser	His	Gln	Ala	Leu	Leu
			340					345					350		
Gln	His	Leu	Leu	Leu	Lys	Glu	Gln	Met	Arg	Gln	Gln	Lys	Leu	Leu	Val
		355					360					365			
Ala	Gly	Gly	Val	Pro	Leu	His	Pro	Gln	Ser	Pro	Leu	Ala	Thr	Lys	Glu
	370					375					380				
Arg	Ile	Ser	Pro	Gly	Ile	Arg	Gly	Thr	His	Lys	Leu	Pro	Arg	His	Arg
385					390					395					400
Pro	Leu	Asn	Arg	Thr	Gln	Ser	Ala	Pro	Leu	Pro	Gln	Ser	Thr	Leu	Ala
				405					410					415	
Gln	Leu	Val	Ile	Gln	Gln	Gln	His	Gln	Gln	Phe	Leu	Glu	Lys	Gln	Lys
			420					425					430		
Gln	Tyr	Gln	Gln	Gln	Ile	His	Met	Asn	Lys	Leu	Leu	Ser	Lys	Ser	Ile
		435					440					445			
Glu	Gln	Leu	Lys	Gln	Pro	Gly	Ser	His	Leu	Glu	Glu	Ala	Glu	Glu	Glu
	450					455					460				
Leu	Gln	Gly	Asp	Gln	Ala	Met	Gln	Glu	Asp	Arg	Ala	Pro	Ser	Ser	Gly
465					470					475					480
Asn	Ser	Thr	Arg	Ser	Asp	Ser	Ser	Ala	Cys	Val	Asp	Asp	Thr	Leu	Gly
				485					490					495	
Gln	Val	Gly	Ala	Val	Lys	Val	Lys	Glu	Glu	Pro	Val	Asp	Ser	Asp	Glu
			500					505					510		
Asp	Ala	Gln	Ile	Gln	Glu	Met	Glu	Ser	Gly	Glu	Gln	Ala	Ala	Phe	Met
		515					520					525			
Gln	Gln	Pro	Phe	Leu	Glu	Pro	Thr	His	Thr	Arg	Ala	Leu	Ser	Val	Arg
	530					535					540				
Gln	Ala	Pro	Leu	Ala	Ala	Val	Gly	Met	Asp	Gly	Leu	Glu	Lys	His	Arg
545					550					555					560
Leu	Val	Ser	Arg	Thr	His	Ser	Ser	Pro	Ala	Ala	Ser	Val	Leu	Pro	His
				565					570					575	
Pro	Ala	Met	Asp	Arg	Pro	Leu	Gln	Pro	Gly	Ser	Ala	Thr	Gly	Ile	Ala
			580					585					590		
Tyr	Asp	Pro	Leu	Met	Leu	Lys	His	Gln	Cys	Val	Cys	Gly	Asn	Ser	Thr
		595					600					605			
Thr	His	Pro	Glu	His	Ala	Gly	Arg	Ile	Gln	Ser	Ile	Trp	Ser	Arg	Leu
	610					615					620				

13/25

Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys  
 625 630 635 640  
 Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu  
 645 650 655  
 Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile  
 660 665 670  
 Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly  
 675 680 685  
 Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser  
 690 695 700  
 Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys  
 705 710 715 720  
 Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro  
 725 730 735  
 Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn  
 740 745 750  
 Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser  
 755 760 765  
 Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln  
 770 775 780  
 Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg  
 785 790 795 800  
 Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val  
 805 810 815  
 Arg Phe Ile Ser Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn  
 820 825 830  
 Cys Ile Ala  
 835

&lt;210&gt; 9

&lt;211&gt; 1791

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 9

ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttccgtgtggg cctggagccc atctcacctt tagacctagg gacagacctc 240  
 aggatgatga tgcccgtggt ggaccctggt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgagca gaacagcagc ttccctctct cagaggcaaa 540  
 gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggta cagggtgccc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaag 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggtctctgt 840  
 ccagttcac caaacaatgg gccaaactgga agtggttactg aaaatgagac ttcggttttg 900  
 cccctacccc ctcatgccga gcaaatggtt tcacagcaac gcatttcta tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggcct 1020  
 cccgcagtgc catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080  
 acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccggca 1140  
 tcttccagcc accctcatgt tacttttagag ggaaagccac ccaacagcag ccaccaggct 1200  
 ctctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260  
 ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320  
 agaggtaccc acaaatggcc ccgtcacaga cccctgaacc gaaccagtc tgcaccttg 1380  
 cctcagagca cgttggctca gctggtcatt caacagcaac accagcaatt cttggagaag 1440  
 cagaagcaat accagcagca gatccacatg aacaaactgc ttctgaaatc tattgaacaa 1500  
 ctgaagcaac caggcagtca ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560

14/25

```

atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgtttgt 1620
gtggatgaca cactgggaca agttggggct gtgaaggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
gtaataggca aagatttagc tccaggattt gtaattaaag tcattatctg a 1791

```

&lt;210&gt; 10

&lt;211&gt; 546

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1      5      10      15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20     25     30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35     40     45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50     55     60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65     70     75     80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Ala Ile Lys Gln Gln
85     90     95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100    105    110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115    120    125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130    135    140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145    150    155    160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165    170    175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180    185    190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195    200    205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
210    215    220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225    230    235    240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
245    250    255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
260    265    270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
275    280    285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
290    295    300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305    310    315    320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
325    330    335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
340    345    350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
355    360    365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
370    375    380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385    390    395    400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala

```

15/25

```

          405          410          415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
          420          425          430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
          435          440          445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
          450          455          460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465          470          475          480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
          485          490          495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
          500          505          510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
          515          520          525
Gln Gln Val Ile Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val
          530          535          540
Ile Ile
545

```

<210> 11  
 <211> 590  
 <212> PRT  
 <213> Homo sapiens

```

<400> 11
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1          5          10          15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
          20          25          30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
          35          40          45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
          50          55          60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65          70          75          80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
          85          90          95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
          100          105          110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
          115          120          125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
          130          135          140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145          150          155          160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
          165          170          175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
          180          185          190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
          195          200          205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
          210          215          220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225          230          235          240
Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
          245          250          255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
          260          265          270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu

```

16/25

```

      275      280      285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val
 290      295      300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser
305      310      315      320
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala
      325      330      335
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys
      340      345      350
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr
      355      360      365
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu
      370      375      380
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu
385      390      395      400
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val
      405      410      415
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro
      420      425      430
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg
      435      440      445
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile
      450      455      460
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln
465      470      475      480
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys
      485      490      495
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Leu Gln Gly Asp
      500      505      510
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg
      515      520      525
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala
530      535      540
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile
545      550      555      560
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Val Ile
      565      570      575
Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val Ile Ile
580      585      590

```

<210> 12  
 <211> 1084  
 <212> PRT  
 <213> Homo sapiens

```

<400> 12
Met Ser Ser Gln Ser His Pro Asp Gly Leu Ser Gly Arg Asp Gln Pro
 1      5      10      15
Val Glu Leu Leu Asn Pro Ala Arg Val Asn His Met Pro Ser Thr Val
      20      25      30
Asp Val Ala Thr Ala Leu Pro Leu Gln Val Ala Pro Ser Ala Val Pro
      35      40      45
Met Asp Leu Arg Leu Asp His Gln Phe Ser Leu Pro Val Ala Glu Pro
50      55      60
Ala Leu Arg Glu Gln Gln Leu Gln Gln Glu Leu Leu Ala Leu Lys Gln
65      70      75      80
Lys Gln Gln Ile Gln Arg Gln Ile Leu Ile Ala Glu Phe Gln Arg Gln
      85      90      95
His Glu Gln Leu Ser Arg Gln His Glu Ala Gln Leu His Glu His Ile
      100      105      110
Lys Gln Gln Gln Glu Met Leu Ala Met Lys His Gln Gln Glu Leu Leu

```



	115						120						125					
Glu His Gln Arg Lys Leu Glu Arg His Arg'Gln Glu Gln Glu Leu Glu																		
Lys 130							135						140					
Lys 145	Gln His Arg Glu Gln Lys Leu Gln Gln Leu Lys Asn Lys Glu Lys																	
Gly	Lys Glu Ser Ala Val Ala Ser Thr Glu Val Lys Met Lys Leu Gln																	
	165						170						175					
Glu Phe Val Leu Asn Lys Lys Lys Ala Leu Ala His Arg Asn Leu Asn																		
	180						185						190					
His Cys Ile Ser Ser Asp Pro Arg Tyr Trp Tyr Gly Lys Thr Gln His																		
	195						200						205					
Ser Ser Leu Asp Gln Ser Ser Pro Pro Gln Ser Gly Val Ser Thr Ser																		
Tyr 210							215						220					
Asn 225	His Pro Val Leu Gly Met Tyr Asp Ala Lys Asp Asp Phe Pro																	
Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu Lys Leu Arg Ser Arg Leu																		
	245						250						255					
Lys Gln Lys Val Ala Glu Arg Arg Ser Ser Pro Leu Leu Arg Arg Lys																		
	260						265						270					
Asp Gly Pro Val Val Thr Ala Leu Lys Lys Arg Pro Leu Asp Val Thr																		
	275						280						285					
Asp Ser Ala Cys Ser Ser Ala Pro Gly Ser Gly Pro Ser Ser Pro Asn																		
	290						295						300					
Asn 305	Ser Ser Gly Ser Val Ser Ala Glu Asn Gly Ile Ala Pro Ala Val																	
Pro Ser Ile Pro Ala Glu Thr Ser Leu Ala His Arg Leu Val Ala Arg																		
	325						330						335					
Glu Gly Ser Ala Ala Pro Leu Pro Leu Tyr Thr Ser Pro Ser Leu Pro																		
	340						345						350					
Asn Ile Thr Leu Gly Leu Pro Ala Thr Gly Pro Ser Ala Gly Thr Ala																		
	355						360						365					
Gly Gln Gln Asp Thr Glu Arg Leu Thr Leu Pro Ala Leu Gln Gln Arg																		
	370						375						380					
Leu Ser Leu Phe Pro Gly Thr His Leu Thr Pro Tyr Leu Ser Thr Ser																		
Pro Leu Glu Arg Asp Gly Gly Ala Ala His Ser Pro Leu Leu Gln His																		
	405						410						415					
Met Val Leu Leu Glu Gln Pro Pro Ala Gln Ala Pro Leu Val Thr Gly																		
	420						425						430					
Leu Gly Ala Leu Pro Leu His Ala Gln Ser Leu Val Gly Ala Asp Arg																		
	435						440						445					
Val Ser Pro Ser Ile His Lys Leu Arg Gln His Arg Pro Leu Gly Arg																		
	450						455						460					
Thr Gln Ser Ala Pro Leu Pro Gln Asn Ala Gln Ala Leu Gln His Leu																		
Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys His Lys Gln Gln																		
	485						490						495					
Phe Gln Gln Gln Gln Leu Gln Met Asn Lys Ile Ile Pro Lys Pro Ser																		
	500						505						510					
Glu Pro Ala Arg Gln Pro Glu Ser His Pro Glu Glu Thr Glu Glu Glu																		
	515						520						525					
Leu Arg Glu His Gln Ala Leu Leu Asp Glu Pro Tyr Leu Asp Arg Leu																		
	530						535						540					
Pro Gly Gln Lys Glu Ala His Ala Gln Ala Gly Val Gln Val Lys Gln																		
	545						550						555					
Glu Pro Ile Glu Ser Asp Glu Glu Glu Ala Glu Pro Pro Arg Glu Val																		
	565						570						575					
Glu Pro Gly Gln Arg Gln Pro Ser Glu Gln Glu Leu Leu Phe Arg Gln																		
	580						585						590					
Gln Ala Leu Leu Glu Gln Gln Arg Ile His Gln Leu Arg Asn Tyr																		
	595						600						605					
Gln Ala Ser Met Glu Ala Ala Gly Ile Pro Val Ser Phe Gly Gly His																		

18/25

610	615	620
Arg Pro Leu Ser Arg	Ala Gln Ser Ser Pro	Ala Ser Ala Thr Phe Pro
625	630	635
Val Ser Val Gln Glu Pro	Pro Thr Lys Pro Arg	Phe Thr Thr Gly Leu
645	650	655
Val Tyr Asp Thr Leu Met	Leu Lys His Gln Cys Thr	Cys Gly Ser Ser
660	665	670
Ser Ser His Pro Glu His	Ala Gly Arg Ile Gln Ser	Ile Trp Ser Arg
675	680	685
Leu Gln Glu Thr Gly Leu	Arg Gly Lys Cys Glu Cys	Ile Arg Gly Arg
690	695	700
Lys Ala Thr Leu Glu Glu	Leu Gln Thr Val His Ser	Glu Ala His Thr
705	710	715
Leu Leu Tyr Gly Thr Asn	Pro Leu Asn Arg Gln Lys	Leu Asp Ser Lys
725	730	735
Lys Leu Leu Gly Ser Leu	Ala Ser Val Phe Val Arg	Leu Pro Cys Gly
740	745	750
Gly Val Gly Val Asp Ser	Asp Thr Ile Trp Asn Glu	Val His Ser Ala
755	760	765
Gly Ala Ala Arg Leu Ala	Val Gly Cys Val Val Glu	Leu Val Phe Lys
770	775	780
Val Ala Thr Gly Glu Leu	Lys Asn Gly Phe Ala Val	Val Arg Pro Pro
785	790	795
Gly His His Ala Glu Glu	Ser Thr Pro Met Gly Phe	Cys Tyr Phe Asn
805	810	815
Ser Val Ala Val Ala Ala	Lys Leu Leu Gln Gln Arg	Leu Ser Val Ser
820	825	830
Lys Ile Leu Ile Val Asp	Trp Asp Val His His Gly	Asn Gly Thr Gln
835	840	845
Gln Ala Phe Tyr Ser Asp	Pro Ser Val Leu Tyr Met	Ser Leu His Arg
850	855	860
Tyr Asp Asp Gly Asn Phe	Pro Gly Ser Gly Ala Pro	Asp Glu Val
865	870	875
Gly Thr Gly Pro Gly Val	Gly Phe Asn Val Asn Met	Ala Phe Thr Gly
885	890	895
Gly Leu Asp Pro Pro Met	Gly Asp Ala Glu Tyr Leu	Ala Ala Phe Arg
900	905	910
Thr Val Val Met Pro Ile	Ala Ser Glu Phe Ala Pro	Asp Val Val Leu
915	920	925
Val Ser Ser Gly Phe Asp	Ala Val Glu Gly His Pro	Thr Pro Leu Gly
930	935	940
Gly Tyr Asn Leu Ser Ala	Arg Cys Phe Gly Tyr Leu	Thr Lys Gln Leu
945	950	955
Met Gly Leu Ala Gly Gly	Arg Ile Val Leu Ala Leu	Glu Gly Gly His
965	970	975
Asp Leu Thr Ala Ile Cys	Asp Ala Ser Glu Ala Cys	Val Ser Ala Leu
980	985	990
Leu Gly Asn Glu Leu Asp	Pro Leu Pro Glu Lys Val	Leu Gln Gln Arg
995	1000	1005
Pro Asn Ala Asn Ala Val	Arg Ser Met Glu Lys Val	Met Glu Ile His
1010	1015	1020
Ser Lys Tyr Trp Arg Cys	Leu Gln Arg Thr Thr Ser	Thr Ala Gly Arg
1025	1030	1035
Ser Leu Ile Glu Ala Gln	Thr Cys Glu Asn Glu Glu	Ala Glu Thr Val
1045	1050	1055
Thr Ala Met Ala Ser Leu	Ser Val Gly Val Lys Pro	Ala Glu Lys Arg
1060	1065	1070
Pro Asp Glu Glu Pro Met	Glu Glu Pro Pro Leu	
1075	1080	

19/25

&lt;211&gt; 3550

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 13

ggggaagaga	ggcacagaca	cagataggag	aagggcaccg	gctggagcca	cttgcaggac	60
tgagggtttt	tgcaacaaaa	ccctagcagc	ctgaagaact	ctaagccaga	tggggtggct	120
ggacgagagc	agctcttggc	tcagcaaaga	atgcacagta	tgatcagctc	agtggatgtg	180
aagtccagaag	ttcctgtggg	cctggagccc	atctcacctt	tagacctaa	gacagacctc	240
aggatgatga	tgcccggtgt	ggaccctgtt	gtccgtgaga	agcaattgca	gcaggaatta	300
cttcttatcc	agcagcagca	acaaatccag	aagcagcttc	tgatagcaga	gtttcagaaa	360
cagcatgaga	acttgacacg	gcagcaccag	gctcagcttc	aggagcatat	caaggaactt	420
ctagccataa	aacagcaaca	agaactccta	gaaaaggagc	agaaactgga	gcagcagagg	480
caagaacagg	aagtagagag	gcctcgcaga	gaacagcagc	ttcctcctct	cagaggcaaa	540
gatagaggac	gagaaagggc	agtggcaagt	acagaagtaa	agcagaagct	tcaagagttc	600
ctactgagta	aatcagcaac	gaaagacact	ccaactaatg	gaaaaaatca	ttccgtgagc	660
cgccatccca	agctctggta	cacggctgcc	caccacacat	cattggatca	aagctctcca	720
ccccttagtg	gaacatctcc	atcctacaag	tacacattac	caggagcaca	agatgcaaa	780
gatgatttcc	cccttcgaaa	aactgcctct	gagcccaact	tgaagggtcg	gtccaggtta	840
aaacagaaag	tggcagagag	gagaagcagc	cccttactca	ggcggaaagga	tggaaatgtt	900
gtcacttcat	tcaagaagcg	aatgtttgag	gtgacagaat	cctcagtcag	tagcagttct	960
ccaggctctg	gtcccagttc	accaaacaat	gggccaactg	gaagtgttac	tgaaaatgag	1020
acttcgggtt	tgccccctac	ccctcatgcc	gagcaaatgg	tttcacagca	acgcattcta	1080
attcatgaag	attccatgaa	cctgctaagt	ctttatacct	ctccttcttt	gccaacattt	1140
accttggggc	ttcccgagct	gccatcccag	ctcaatgctt	cgaattcact	caaagaaaag	1200
cagaagtgtg	agacgcagac	gcttaggcaa	gggtgttctc	tgcttgggca	gtatggaggc	1260
agcatcccgg	catcttccag	ccaccctcat	gttactttag	agggaaagcc	acccaacagc	1320
agccaccagg	ctctcctgca	gcatttatta	ttgaaagaac	aaatgcgaca	gcaaaagctt	1380
cttgtagctg	gtggagttcc	cttacatcct	cagtctccct	tggcaacaaa	agagagaatt	1440
tcacctggca	ttagagggtac	ccacaaattg	ccccgtcaca	gacccctgaa	ccgaaccag	1500
tctgcacctt	tgctcagag	cacgttggct	cagctgggtc	ttcaacagca	acaccagcaa	1560
ttcttggaga	agcagaagca	ataccagcag	cagatccaca	tgaacaaact	gctttcgaaa	1620
tctattgaac	aactgaagca	accaggcagt	caccttgagg	aagcagagga	agagcttcag	1680
ggggaccagg	cagtcagga	agacagagcg	ccctctagt	gcaacagcac	taggagcgag	1740
agcagtgcct	gtgtggatga	cacactggga	caagttgggg	ctgtgaaggt	caaggaggaa	1800
ccagtggaca	gtgatgaaga	tgctcagatc	caggaaatgg	aatctgggga	gcaggctgct	1860
tttatgcaac	aggtaatatg	caaagattta	gctccaggat	ttgtaatata	agtcattatc	1920
tgacctttcc	tggaaaccac	gcacacacgt	gcgtctctct	tgcgccaagc	tccgctggct	1980
gcggttggca	tggatggatt	agagaaacac	cgctctgctc	ccaggactca	ctcttccctt	2040
gctgcctctg	ttttacctca	cccagcaatg	gaccgcccc	tccagcctgg	ctctgcaact	2100
ggaattgcct	atgacctctt	gatgctgaaa	caccagtgcg	tttgtggcaa	ttccaccacc	2160
caccctgagc	atgctggacg	aatacagagt	atctggtcac	gactgcaaga	aactgggctg	2220
ctaaataaat	gtgagcgaat	tcaaggtcga	aaagccagcc	tggaggaaat	acagcttggt	2280
cattctgaac	atcactcact	gttgtatggc	accaaccccc	tggacggaca	gaagctggag	2340
cccaggatac	tcctaggtga	tgactctcaa	aagttttttt	cctcattacc	ttgtgtggga	2400
cttgggggtg	acagtgcac	catttggaat	gagctacact	cgctccgtgc	tgcacgcacg	2460
gctgttggct	gtgtcatcga	gctggcttcc	aaagtggcct	caggagagct	gaagaatggg	2520
tttgctgttg	tgaggccccc	tggccatcac	gctgaagaat	ccacagccat	gggttctctg	2580
ttttttaatt	cagttgcaat	taccgcaaaa	tacttgagag	accaactaaa	tataagcaag	2640
atattgattg	tagatctgga	tgttcaccat	ggaaacggta	cccagcaggc	cttttatgct	2700
gaccccagca	tcctgtacat	ttcactccat	cgctatgatg	aagggaactt	tttccctggc	2760
agtggagccc	caaatgaggt	tcggtttatt	tccttagagc	cccactttta	tttgtatctt	2820
tcaggtaatt	gcattgcatg	attaccctta	attttcttgt	cctttgctgg	tgttttaaat	2880
tacacgagat	tactgaattg	tcccatggga	ccaagaacca	gtgcagaaca	agtgcataac	2940
ccagagcact	gtttgtcagg	gaaggttggg	ctgatttgat	gtgtgttttg	atgtttatct	3000
caagagctcc	catgtgcttg	ttttctctct	ttcttgcttt	cttccatttg	ctctctctct	3060
tgccaccgtg	ggtgtgtctt	tctcttccca	ggttggaaca	ggccttggag	aagggtacaa	3120
tataaatatt	gcctggacag	gtggccttga	tcctcccatg	ggagatgttg	agtaccttga	3180
agcattcagg	accatcgtga	agcctgtggc	caaagagttt	gatccagaca	tggtcttagt	3240
atctgttgga	tttgatgcac	tggaaaggcca	caccctctct	ctaggagggg	acaaagtgc	3300
ggcaaaatgt	tttggctcatt	tgacgaagca	atgatgaca	ttggctgatg	gacgtgtggg	3360
gttggctcta	gaaggaggac	atgatctcac	agccatctgt	gatgcacacg	aagcctgtgt	3420

20/25

aatgcccctt ctaggaaatg agctggagcc acttgacagaa gatatttctcc accaaagccc 3480  
gaatatgaat gctgttattt ctttacagaa gatcattgaa attcaaagta tgtcttttaa 3540  
gttctcttaa 3550

&lt;210&gt; 14

&lt;211&gt; 7699

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 14

cccattcgcc attcaggctg cgcaactggt gggaagggcg atcgggtgagg gcctcttcgc 60  
tattacgccca gctggcgaaa ggggatgtg ctgcaaggcg attaggttg gtaacgcccc 120  
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180  
ttggccatta gccatattat tcattgggta tatagcataa atcaatattg gctattggcc 240  
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300  
accgcatatg tgacattgat tattgactag ttattaatag taatcaatta cgggggtcatt 360  
agtcatagc ccatatatgg agttccgctg tacataaact acggtaaatg gcccgccctg 420  
cgaccgcccc gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480  
ccaataggga cttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540  
gcagtcacac aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa 600  
tggcccgccct agcattatgc ccagtacatg acctacggg agtttccctac ttggcagtac 660  
atctacgtat tagtcatcgc tattaccatg gtgatgcggg tttggcagta caccaatggg 720  
cgtggatagc gggttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780  
agtttgtttt ggcacaaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840  
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900  
gtgaaccgctc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960  
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggccc tggagcccat 1020  
ctcaccttta gacctaaagg cagacctcag gatgatgatg cccgtgggtgg accctgttgt 1080  
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140  
gcagcttctg atagcagagt ttcagaaaca gcatgagaac ttgacacggc agcaccaggc 1200  
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260  
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgagaga 1320  
acagcagctt cctcctctca gaggcaaaga tagaggacga gaaagggcag tggcaagtac 1380  
agaagtaaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440  
aactaatgga aaaaatcatt ccgtgagccg ccattcccaag ctctggtaca cggctgcccc 1500  
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560  
cacattacca ggagcacaaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620  
gcccacttg aaggtgcggg ccagggttaa acagaaagtg gcagagagga gaagcagccc 1680  
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740  
gacagaatcc tcagtcagta gcagttctcc aggtctctgg cccagttcac caacaatgg 1800  
gccaactgga agtggtactg aaaatgagac ttcggttttg cccctaccct ctcagccga 1860  
gcaaatggtt tcacagcaac gcattctaatt tcatgaagat tccatgaacc tgctaagtct 1920  
ttatacctct ccttctttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980  
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040  
tgttccctctg cctgggcagt atggaggcag catcccggca tcttccagcc accctcatgt 2100  
tacttttagag ggaaagccac ccaacagcag ccaccaggct ctcctgcagc atttattatt 2160  
gaaagaacaa atgcgacagc aaaagcttct tgtagctggg ggagttccct tacatcctca 2220  
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtacc ccaaatggcc 2280  
ccgtcacaga cccctgaacc gaacccagtc tgacaccttg cctcagagca cgttgggtca 2340  
gctggctcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400  
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtc 2460  
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520  
ctctagtggc aacagcacta ggagcgacag cagtgttgt gtggatgaca cactgggaca 2580  
agttggggct gtgaagggtc agggaggaacc agtgagcagt gatgaagat ctcagatcca 2640  
ggaaatggaa tctggggagc aggtgtcttt tatgcaacag cctttctctg aaccacgca 2700  
cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg gttggcatgg atggattaga 2760  
gaaacaccgt ctcgtctcca ggactcactc ttcccctgct gcctctgttt tacctcacc 2820  
agcaatggac cgccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880  
gctgaaacac cagtgcgttt gtggcaattc caccaccac cctgagcatg ctggacgaat 2940  
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000  
aggctgaaaa gccagcctgg aggaaatata gcttgttcat tctgaacatc actcactgtt 3060  
gtatggcacc aacccccctg acggacagaa gctggacccc aggtactcc taggtgatga 3120

ctctcaaaag	tttttttccct	cattaccttg	tgggtggactt	ggggtggaca	gtgacacccat	3180
ttggaatgag	ctacactcgt	ccggtgctgc	acgcatggct	gttggctgtg	tcacgcagct	3240
ggcttccaaa	gtggcctcag	gagagctgaa	gaatgggttt	gctgttgtga	ggccccctgg	3300
ccatcacgct	gaagaatcca	cagccatggg	gttctgcttt	tttaattcag	ttgcaattac	3360
cgccaaatac	ttgagagacc	aactaaatat	aagcaagata	ttgattgtag	atctggatgt	3420
tcaccatgga	aacggtaccc	agcaggcctt	ttatgctgac	cccagcatcc	tgtacatttc	3480
actccatcgc	tatgatgaag	ggaacttttt	ccctggcagt	ggagcccca	atgaggttgg	3540
aacaggcctt	ggagaagggt	acaatataaa	tattgcctgg	acagggtggc	ttgatcctcc	3600
catgggagat	gttgagtacc	ttgaagcatt	caggaccatc	gtgaagcctg	tggccaaaga	3660
gtttgatcca	gacatggtct	tagtatctgc	tggatttgat	gcattggaag	gccacacccc	3720
tcctctagga	gggtacaaag	tgacggcaaa	atgttttggg	catttgacga	agcaattgat	3780
gacattgggt	gatggacgtg	tgggtgttggc	tctagaagga	ggacatgatc	tcacagccat	3840
ctgtgatgca	tcagaagcct	gtgtaaatac	ccttctagga	aatgagctgg	agccacttgc	3900
agaagatatt	ctccacaaa	gcccgaatat	gaatgctgtt	atttctttac	agaagatcat	3960
tgaatttcaa	agtatgtcct	taaagttctc	tggatccggg	accagattac	aaggacgacg	4020
atgacaagta	gatcccggtg	ggcatccctg	tgaccctcc	ccagtgcctc	tcctggcctt	4080
ggaagttggc	actccagtcg	ccaccagcct	tgctcctaata	aaattaagtt	gcacattttt	4140
gtctgactag	gtgtcctcta	taatattatg	gggtggaggg	gggtgggatg	gagcaagggg	4200
cccaagtggg	gaagacaacc	tgtagggcct	gcgggttcta	ttcgggaacc	aagctggagt	4260
gcagtggcac	aatcttggct	cactgcaatc	tcgcctcctt	gggttcaagc	gattctcctg	4320
cctcagcctc	ccgagttgtt	gggattccag	gcattgcata	ccaggctcag	ctaatttttg	4380
tttttttggg	agagacgggg	tttcaccata	ttggccaggc	tgggtctcaa	ctcctaattc	4440
caggtgatct	accacacttg	gcctcccaaa	ttgctgggat	tacaggcggt	aaccactgct	4500
cccttccctg	tccttctgat	tttaaaataa	ctataccagc	aggaggacgt	ccagacacag	4560
cataggctac	ctgccatggc	ccaaccgggt	ggacatttga	gttgcttggc	tggcactgtc	4620
ctctcatgcg	ttgggtccac	tcagtagatg	cctgttgaat	tgggtacgcg	gccagcttct	4680
gtggaatgtg	tgtcagttag	gggtgtgaaa	gtccccaggc	ttcccagcag	gcagaagtat	4740
gcaaagcatg	catctcaatt	agtcagcaac	caggtgtgga	aaagtcccca	ggctccccag	4800
caggcagaag	tatgcaaagc	atgcatctca	attagttagc	aaccatagtc	ccgcccctaa	4860
ctccgcccac	ccgcccctta	actccgcccc	gttccgcccc	ttctccgccc	catggctgac	4920
taattttttt	tatttatgca	gaggccgagg	ccgcctcgcc	ctctgagcta	ttccagaagt	4980
agtgaggagg	cttttttggg	ggcctaggct	tttgcaaaaa	gtcctctcag	gaactgaaaa	5040
accagaaagt	taattcccta	tagtgagtcg	tattaaattc	gtaatcatgg	tcataagctg	5100
ttcctgtgtg	aaattgttat	ccgctcacia	ttccacacaa	catacgagcc	ggaagcataa	5160
agtgtaaaag	ctgggggtgc	taatgagtga	gctaactcac	attaattgcg	ttgcgctcac	5220
tgcccgcctt	ccagtcggga	aacctgtcgt	gccagctgca	ttaatgaatc	ggccaacgcg	5280
cggggagagg	cggtttgctg	attggggcgt	cttccgcttc	ctcgtcact	gactcgctgc	5340
gctcggctcg	tcggctgcgg	cgagcgggat	cagctcactc	aaaggcggtg	atacgggttat	5400
ccacagaatc	aggggataac	gcaggaaaag	acatgtgagc	aaaaggccag	caaaaggcca	5460
ggaaccgtaa	aaaggccgcg	ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	5520
atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	gacaggacta	taaaagatac	5580
aggcgtttcc	ccctggaagc	tcctctcgtc	gctctcctgt	tcgaccctg	ccgcttaccg	5640
gatacctgtc	cgcttttctc	ccttcgggaa	gcgtggcgct	ttctcaatgc	tcacgctgta	5700
ggtatctcag	ttcgggtgtag	gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	5760
ttcagcccg	ccgctgcgcc	ttatccggta	actatcgtct	tgagtcacac	ccggtaaagc	5820
acgacttatc	gccactggca	gcagccactg	gtaacaggat	tagcagagcg	aggatgttag	5880
gcggtgtctc	agagttcttg	aagtgggtggc	ctaactacgg	ctacactaga	agaacagtat	5940
ttggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	aagagttggg	agctcttgat	6000
ccggcaaaaa	aaccaccgct	ggtagcgggt	gtttttttgt	ttgcaagcag	cagattacgc	6060
gcagaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	tacgggggtc	gacgctcagt	6120
ggaacgaaaa	ctcacgttaa	gggatttttg	tcatgagatt	atcaaaaagg	atcttcacct	6180
agatcctttt	aaattaaaaa	tgaagtttta	aatcaatcta	aagtataat	gagtaaaact	6240
ggtctgacag	ttaccaatgc	ttaatcagtg	aggcacctat	ctcagcgatc	tgtctatttc	6300
gttcatccat	agttgcctga	ctccccgtcg	tgtagataac	tacgatacgg	gagggettac	6360
catctggccc	cagtgtgca	atgataccgc	gagacccacg	ctcaccggct	ccagatttat	6420
cagcaataaa	ccagccagcc	ggaaggcccg	agcgagaaag	tggctcctga	actttatccg	6480
cctccatcca	gtctattaat	tgttgccggg	aagctagagt	aagtagttcg	ccagtttaata	6540
gtttgcgcaa	cgttgttgcc	attgctacag	gcacgtgtgt	gtcacgctcg	tcgtttggta	6600
tggcttcatt	cagctccggg	tcccaacgat	caaggcgagt	tacatgatcc	cccatgttgt	6660
gcaaaaaagc	ggttgactcc	ttcggctctc	cgatcgttgt	cagaagtaag	ttggcccgag	6720
tgttatcact	catgtgtatg	gcagcactgc	ataattctct	tactgtcatg	ccatccgtaa	6780
gatgcttttc	tgtgactggg	gagtactcaa	ccaagtcatt	ctgagaatag	tgtatgcccc	6840

22/25

gaccgagttg ctcttgcccg gcgtcaatac gggataatac cgcgccacat agcagaactt 6900  
taaaagtgt catcattgga aaacgttctt cggggcgaaa actctcaagg atcttaccgc 6960  
tggtgagatc cagttcgatg taaccactc gtgcaccaa ctgatcttca gcatctttta 7020  
ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa 7080  
taagggcgac acggaaatgt tgaatactca tactcttctt ttttcaatat tattgaagca 7140  
tttatcaggg ttattgtctc atgagcggat acatatttga atgtatttag aaaaataaac 7200  
aaataggggt tccgcgcaca tttccccgaa aagtgccacc tgacgcgccc tgtagcggcg 7260  
cattaagcgc ggcggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcgccc 7320  
tagcgccgc tcctttcgtc ttcttccctt cctttctcgc cacgttcgcc ggctttcccc 7380  
gtcaagctct aaatcggggc atccctttag ggttccgatt tagtgcttta cggcacctcg 7440  
accccaaaaa acttgattag ggtgatgggt cacgtagtgg gccatcgccc tgatagacgg 7500  
tttttcgccc tttgacgttg gagtccacgt tctttaatag tggactcttg ttccaaactg 7560  
gaacaacact caaccctatc tcggtctatt cttttgattt ataagggaatt ttgccgattt 7620  
cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat ttttaacaaa 7680  
tattaaacgt ttacaattt 7699

&lt;210&gt; 15

&lt;211&gt; 7303

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

ccatttcgcc attcaggctg cgcaactggt ggggaagggcg atcgggtgcgg gcctcttcgc 60  
tattacgccca gctggcgaaa gggggatgtg ctgcaaggcg attaagttgg gtaacgcccc 120  
gggttttccc agtcacgacg ttgtaaaaac acggccagtg ccaagctgat ctaatcaata 180  
ttggccatta gccatattat tcatttggtta tatagcataa atcaatattg gctattggcc 240  
attgcatacg ttgtatccat atcataatat gtacatttat attgggtcat gtccaacatt 300  
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360  
agttcatagc ccataatatg agttccgcgt tacataactt acggtaaatg gccgccttg 420  
cgaccgcccc gcgacccccg cccgttgacg tcaatagtga cgtatgttcc catagtaacg 480  
ccaatagggc ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540  
gcagtacatc aagtgtatca tatgccagt cgcgcccta ttgacgtcaa tgacggtaaa 600  
tgcccgccct agcattatgc ccagtacatg acctacggg agtttctac ttggcagtac 660  
atctacgtat tagtcatcgc tattaccatg gtgatgcggg tttggcagta caccaatggg 720  
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg 780  
agtttgtttt ggcacaaaaa tcaacgggac tttccaaaat gtcgtaataa ccccgccccg 840  
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900  
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960  
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggcc tggagcccat 1020  
ctcaccttta gacctaaagg cagacctcag gatgatgatg cccgtgggtg accctgttgt 1080  
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140  
tcagcttctg atagcagagt ttcagaaaac gcatgagaac ttgacacggc agcaccaggc 1200  
gcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260  
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgagaga 1320  
acagcagctt cctcctctca gaggcaaaag tagaggacga gaaagggcag tggcaagtac 1380  
agaagtaaa cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440  
aactaatgga aaaaatcatt ccgtgagccg ccatcccaag ctctggtaca cggctgcccc 1500  
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560  
cacattacca ggagcacaag atgcaaagga tgatttcccc cttcgaaaaa ctgcctctga 1620  
gcccaacttg aaggtgcggg ccaggttaaa acagaaaagt gcagagagga gaagcagccc 1680  
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740  
gacagaatcc tcagtcagta gcagttctcc aggtctgtt cccagttcac caaacaattg 1800  
gccaaactgga agtggtactg aaaaatgagac ttcggttttg cccctacccc ctcatgccga 1860  
gcaaattggtt tcacagcaac gcatttcta tcatgaagat tccatgaacc tgctaagtct 1920  
ttatacctct cttcttttgc ccaacattac cttggggctt cccgcagtgc catcccagct 1980  
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040  
tgttctctcg cctgggcagt atggaggcag ccatccggca tcttccagcc accctcatgt 2100  
tacttttagg ggaaagccac ccaacagcag ccaccaggct ctctgcagc atttattatt 2160  
gaaagaacaa atgcgacagc aaaagcttct tgtagctggg ggagttccct tacatcctca 2220  
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280

ccgtcacaga cccctgaacc gaaccagtc tgcacctttg cctcagagca cgttggtctca 2340  
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400  
gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtc 2460  
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520  
ctctagtggc aacagcacta ggagcgacag cagtgtctgt gtggatgaca cactgggaca 2580  
agtgggggct gtgaaggta aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640  
ggaaatggaa tctgggggagc aggtgtcttt tatgcaacag cctttcctgg aaccacgca 2700  
cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg gttggcatgg atggattaga 2760  
gaaacacogt ctctgtctcca ggactcactc ttcccctgct gcctctgttt tacctcacc 2820  
agcaatggac cgccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880  
gctgaaacac cagtgcgttt gtggcaattc caccaccac cctgagcatg ctggacgaat 2940  
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000  
aggtcgaaaa gccagcctgg aggaatata gcttgttcat tctgaacatc actcactgtt 3060  
gtatggcacc aaccctctgg acggacagaa gctggacccc aggatactcc taggtgatga 3120  
ctctcaaaag ttttttctct cattaccttg tgggtgactt ggggtggaca gtgacacat 3180  
ttggaatgag ctacactcgt ccggtgtctgc acgcatggct gttggctgtg tcatcgagct 3240  
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtga ggccccctgg 3300  
ccatcacgct gaagaatcca cagccatggg gttctgcttt ttaattcag ttgcaattac 3360  
cgccaaatc ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420  
tcaccatgga aacggtaccc agcaggcctt ttatgtctgac ccagcatcc tgtacatttc 3480  
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagcccaa atgagggttcg 3540  
gtttatttct ttagagcccc acttttattt gtatctttca ggtaattgca ttgcaggatc 3600  
cgttaccaga ttacaaggac gacgatgaca agtagatccc ggggtggcatc cctgtgacc 3660  
ctccccagt cctctcctgg ccttggaaat tgccactcca gtgcccacca gccttgtcct 3720  
aataaaatta agttgcatca ttttgtctga ctagggtctc tctataatat tatggggtgg 3780  
aggggggtgg tatggagcaa gggggccaag ttgggaagac aacctgtagg gcctgcgggg 3840  
tctattcggg aaccaagctg gagtgcagtg gcacaatctt ggctcactgc aatctccgcc 3900  
tcttgggttc aagcgattct cctgcctcag cctcccagat tgttgggatt ccaggcatgc 3960  
atgaccaggc tcagctaatt tttgtttttt ttgttagagac ggggtttcac catattggcc 4020  
aggctggctt ccaactccta atctcagtg atctaccac cttggcctcc caaattgctg 4080  
ggattacagg cgtgaaccac tgctcccttc cctgtccttc tgattttaaa ataactatac 4140  
cagcaggagg acgtccagac acagcatagg ctacctgcca tggcccaacc ggtgggacat 4200  
ttgagtgtgt tgcttggcac tgtcctctca tgcgttgggt ccactcagta gatgcctgtt 4260  
gaattgggta cgcgccagc ttctgtggaa tgtgtgtcag ttagggtgtg gaaagctccc 4320  
aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag caaccagtg 4380  
tggaaggc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 4440  
cagcaaccat agtcccgcc ctaactccgc ccatcccgcc cctaactccg cccagttccg 4500  
cccattctcc gcccatggc tgactaattt tttttattta tgcagaggcc gaggcgcct 4560  
cggcctctga gctattccag aagtagtgag gaggtttttt tggaggccta ggcttttgca 4620  
aaaagctcct ctaggaactg aaaaaccaga aagtttaatt cctatagtga gtctatttaa 4680  
attcgtaatc atggtcatag ctgtttctct tgtgaaattg ttatccgctc acaattccac 4740  
acaacatacg agccggaagc ataaagtgt aagcctgggg tgcctaata gtgagctaac 4800  
tcacattaat tgcgttgcgc tcaactgccg ctttccagtc gggaaacctg tctgtccagc 4860  
tgcattaatg aatcgcccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg 4920  
cttctcgct cactgactcg ctgcgctcgg tcttccggt gcggcgagcg gtatcagctc 4980  
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 5040  
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cggttgcgt gcgtttttcc 5100  
ataggctccg cccctctgac gagcatcaca aaaatcgagc ctcaagtcag aggtggcgaa 5160  
acccgacagg actataaaga taccaggcgt ttcccctgg aagctccctc gtgcgctctc 5220  
ctgttccgac ctgcgctt accggatacc tctccgctt tctcccttcg ggaagcgtgg 5280  
cgcttctcca atgtcacgc tgtaggtatc tcagttcggg ttaggtcgtt cgctccaagc 5340  
tgggctgtgt gcacgaacc cccgttcagc ccgaccgctg cgcttatcc ggtaactatc 5400  
gtcttgagtc caaccggta agacacgact tatcgccact ggcagcagcc actggtaaca 5460  
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 5520  
acggctacac tagaagaaca gtatttggtg tctgcgctct gctgaagcca gttactctcg 5580  
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt 5640  
ttgtttgcaa gcagcagatt acgcgagaa aaaaaggatc tcaagaagat cctttgatct 5700  
tttctacggg gtctgacgct cagtggaacg aaaactcac ttaagggtatt ttggtcatga 5760  
gattatcaaa aaggatctt acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 5820  
tctaaagtat atatgagtaa acttggctcg acagttaaca atgcttaac agtggagc 5880  
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 5940  
taactacgat acgggagggc ttaccatctg gcccagtgcc tgcaatgata ccgcgagacc 6000

24/25

```

cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 6060
gaagtgggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 6120
gagtaagtag ttcgccagtt aatagtttgc gcaacggtgt tgccattgct acaggcatcg 6180
tggtgtcacg ctcgctggtt ggtagggctt cattcagctc cggttcccaa cgatcaaggc 6240
gagttacatg atcccccatg ttgtgcaaaa aagcgggttag ctcttcggt cctccgatcg 6300
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 6360
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtagtac tcaaccaagt 6420
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgta atacgggata 6480
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 6540
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac 6600
ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 6660
ggcaaaatgc cgcaaaaaag ggaataaggc cgacacggaa atgttgaata ctcatactct 6720
tcctttttca atattattga agcatttata agggttattg tctcatgagc ggatacatat 6780
ttgaatgtat ttagaaaaat aaacaaatag gggttccgag cacatttccc cgaaaagtgc 6840
cacctgacgc gccctgtagc ggcgcattaa gcgcggcggt tgtggtggtt acgcgcagcg 6900
tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttcctttc 6960
tcgccacggt cgccggcttt ccccgtaag ctctaaatcg gggcatccct ttaggggttc 7020
gatttagtgc tttacggcac ctcgacccca aaaaacttga ttaggggtgat ggttcacgta 7080
gtggggccatc gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta 7140
atagtggact cttgttccaa actggaacaa cactcaacct tatctcggtc tattcttttg 7200
atttataagg gattttgccg atttcggcct attgggttaa aaatgagctg atttaacaaa 7260
aatttaacgc gaattttaac aaaatattaa acgtttacaa ttt 7303

```

&lt;210&gt; 16

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 16

ccatggaaac ggtaccagc agg

24

&lt;210&gt; 17

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 17

cactccatcg ctatgatgaa ggg

23

&lt;210&gt; 18

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 18

agttcccttc atcatagcga tgg

23

&lt;210&gt; 19

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence



25/25

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 19

aatgtacagg atgctggggt

20

&lt;210&gt; 20

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 20

cccttgtagc tggtaggtt ccctt

25

&lt;210&gt; 21

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 21

tgtgtcatcg agctggcttc

20

&lt;210&gt; 22

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 22

atcttctgca agtggctcca

20

7/25

Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His Pro Ala Met Asp  
 610 615 620  
 Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala Tyr Asp Pro Leu  
 625 630 635 640  
 Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr Thr His Pro Glu  
 645 650 655  
 His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu Gln Glu Thr Gly  
 660 665 670  
 Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys Ala Ser Leu Glu  
 675 680 685  
 Glu Ile Gln Leu Val His Ser Glu His His Ser Leu Leu Tyr Gly Thr  
 690 695 700  
 Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile Leu Leu Gly Asp  
 705 710 715 720  
 Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly Gly Leu Gly Val  
 725 730 735  
 Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser Gly Ala Arg  
 740 745 750  
 Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys Val Ala Ser Gly  
 755 760 765  
 Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro Gly His His Ala  
 770 775 780  
 Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn Ser Val Ala Ile  
 785 790 795 800  
 Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser Lys Ile Leu Ile  
 805 810 815  
 Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln Gln Ala Phe Tyr  
 820 825 830  
 Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg Tyr Asp Glu Gly  
 835 840 845  
 Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val Arg Phe Ile Ser  
 850 855 860  
 Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn Cys Ile Ala  
 865 870 875

&lt;210&gt; 5

&lt;211&gt; 3054

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 5

ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttctctgtgg cctggagccc atctcacctt tagacctaa gacagacctc 240  
 aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaattccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaaact 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcacgcagca gaacagcagc ttctctctct cagaggcaaa 540  
 gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720  
 cccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaa 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggctctggt 840  
 ccagttcac caaacaatgg gccaactgga agtggtactg aaaatgagac ttccggtttt 900  
 cccctaccc ctcacatgcc gcaaatggtt tcacagcaac gcattctaatt tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020  
 cccgcagtgc catccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080  
 acgcagacgc ttaggcaagg tgttcctctg cctgggcagt atggaggcag catcccgga 1140  
 tcttcagacc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200

8/25

```

ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260.
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaattgcc cgcgcacaga cccctgaacc gaaccagtc tgcacctttg 1380
cctcagagca cgttggtcga gctgggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtc ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgccttg 1620
gtggatgaca cactgggaca agttggggct gtgaaggatc aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggtgccttt tatgcaacag 1740
cctttcctgg aaccacgca cacacgtgcy ctctctgtgc gccaaagctcc gctgggtgcy 1800
gttggcatgg atggattaga gaaacaccgt ctctgtctca ggactcactc tccccctgct 1860
gcctctgttt tacctcacc agcaatggac cgccccctcc agcctggctc tgcaactgga 1920
attgcctatg accccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaaattca aggtcgaaaa gccagcctgg aggaaatata gcttggtcat 2100
tctgaacatc actcactgtt gtatggcacc aacccccctg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaaag tttttttcct cattaccttg tgggtggactt 2220
ggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcctcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggccccctgg ccatacagct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcaccatgga aacggtagcc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccc aaatgaggttg aacaggcctt ggagaagggt acaatataaa tattgcctgg 2640
acaggtggcc ttgatcctcc catgggagat gttgagtacc ttgaagcatt caggaccatc 2700
gtgaagcctg tggccaaaga gtttgatcca gacatggtct tagtatctgc tggatttgat 2760
gcattggaag gccacacccc tcctctagga gggtacaaag tgacggcaaa atgttttgg 2820
catttgacga agcaattgat gacattggct gatggacgtg tgggtgttggc tctagaagga 2880
ggacatgatc tcacagccat ctgtgatgca tcagaagcct gtgtaaagtc ccttctagga 2940
aatgagctgg agccacttgc agaagatatt ctccacaaa gcccgaaat gaatgctgtt 3000
atctctttac agaagatcat tgaaattcaa agtatgtctt taaagttctc ttaa 3054

```

&lt;210&gt; 6

&lt;211&gt; 967

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 6

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1 5 10 15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20 25 30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35 40 45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50 55 60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65 70 75 80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Ala Ile Lys Gln Gln
85 90 95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100 105 110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115 120 125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130 135 140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145 150 155 160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165 170 175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180 185 190

```

9/25

Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp  
 195 200 205  
 Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser  
 210 215 220  
 Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly  
 225 230 235 240  
 Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala  
 245 250 255  
 Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met  
 260 265 270  
 Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu  
 275 280 285  
 Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys  
 290 295 300  
 Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu  
 305 310 315 320  
 Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His  
 325 330 335  
 Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu  
 340 345 350  
 Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val  
 355 360 365  
 Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu  
 370 375 380  
 Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg  
 385 390 395 400  
 Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala  
 405 410 415  
 Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys  
 420 425 430  
 Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile  
 435 440 445  
 Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu  
 450 455 460  
 Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly  
 465 470 475 480  
 Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly  
 485 490 495  
 Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu  
 500 505 510  
 Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met  
 515 520 525  
 Gln Gln Pro Phe Leu Glu Pro Thr His Thr Arg Ala Leu Ser Val Arg  
 530 535 540  
 Gln Ala Pro Leu Ala Ala Val Gly Met Asp Gly Leu Glu Lys His Arg  
 545 550 555 560  
 Leu Val Ser Arg Thr His Ser Ser Pro Ala Ala Ser Val Leu Pro His  
 565 570 575  
 Pro Ala Met Asp Arg Pro Leu Gln Pro Gly Ser Ala Thr Gly Ile Ala  
 580 585 590  
 Tyr Asp Pro Leu Met Leu Lys His Gln Cys Val Cys Gly Asn Ser Thr  
 595 600 605  
 Thr His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg Leu  
 610 615 620  
 Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys  
 625 630 635 640  
 Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu  
 645 650 655  
 Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile  
 660 665 670  
 Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly  
 675 680 685

10/25

Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser  
 690 695 700  
 Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys  
 705 710 715 720  
 Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro  
 725 730 735  
 Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn  
 740 745 750  
 Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser  
 755 760 765  
 Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln  
 770 775 780  
 Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg  
 785 790 795 800  
 Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val  
 805 810 815  
 Gly Thr Gly Leu Gly Glu Gly Tyr Asn Ile Asn Ile Ala Trp Thr Gly  
 820 825 830  
 Gly Leu Asp Pro Pro Met Gly Asp Val Glu Tyr Leu Glu Ala Phe Arg  
 835 840 845  
 Thr Ile Val Lys Pro Val Ala Lys Glu Phe Asp Pro Asp Met Val Leu  
 850 855 860  
 Val Ser Ala Gly Phe Asp Ala Leu Glu Gly His Thr Pro Pro Leu Gly  
 865 870 875 880  
 Gly Tyr Lys Val Thr Ala Lys Cys Phe Gly His Leu Thr Lys Gln Leu  
 885 890 895  
 Met Thr Leu Ala Asp Gly Arg Val Val Leu Ala Leu Glu Gly Gly His  
 900 905 910  
 Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Asn Ala Leu  
 915 920 925  
 Leu Gly Asn Glu Leu Glu Pro Leu Ala Glu Asp Ile Leu His Gln Ser  
 930 935 940  
 Pro Asn Met Asn Ala Val Ile Ser Leu Gln Lys Ile Ile Glu Ile Gln  
 945 950 955 960  
 Ser Met Ser Leu Lys Phe Ser  
 965

&lt;210&gt; 7

&lt;211&gt; 3367

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 7

ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgaggggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtgggatgtg 180  
 aagtcagaag ttctgtggg cctggagccc atctcacctt tagacctaat gacagacctc 240  
 aggatgatga tgcccgtggt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgcaga gaacagcagc ttctctctct cagaggcaaa 540  
 gatagaggac gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggtg cacggctgcc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaa 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggtctctgtg 840  
 ccagttcac caaacaatgg gccaaactgga agtggtactg aaaatgagac ttccggttttg 900  
 cccctaccc ctcattgccg gcaaatggtt tcacagcaac gcattctaatt tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttggggctt 1020  
 cccgcagtcg catcccagct caatgcttcg aattcactca aagaaaagca gaagtgtgag 1080

11/25

```

acgcagacgc ttaggcaagc tgttcctctg cctgggcagt atggaggcag catccccggca 1140
tcttccagcc accctcatgt tacttttagag ggaaagccac ccaacagcag ccaccaggct 1200
ctcctgcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggc 1260
ggagttccct tacatcctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320
agaggtaccc acaaatgtcc ccgtcacaga cccctgaacc gaaccagtc tgcacctttg 1380
cctcagagca cgttgggtca gctgggtcatt caacagcaac accagcaatt cttggagaag 1440
cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500
ctgaagcaac caggcagtc ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg 1560
atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgcctgt 1620
gtggatgaca cactgggaca agttggggct gtgaagggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcatatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
cctttcctgg aaccacgcga cacacgtgcg ctctctgtgc gccaaactcc gctggctgcg 1800
gttggcatgg atggattaga gaaacaccgt ctgctctcca ggactcactc ttcccctgct 1860
gcctctgttt tacctcacc agcaatggac cgccccctcc agcctggctc tgcaactgga 1920
attgcctatg accccttgat gctgaaacac cagtgcgttt gtggcaattc caccacccac 1980
cctgagcatg ctggacgaat acagagtatc tggtcacgac tgcaagaaac tgggctgcta 2040
aataaatgtg agcgaattca aggtcgaaaa gccagcctgg aggaaataca gcttgttcat 2100
tctgaacatc actcactgtt gtatggcacc aacccctgg acggacagaa gctggacccc 2160
aggatactcc taggtgatga ctctcaaaag ttttttctct cattaccttg tgggtggactt 2220
gggggtggaca gtgacaccat ttggaatgag ctacactcgt ccggtgctgc acgcatggct 2280
gttggctgtg tcatcgagct ggcttccaaa gtggcctcag gagagctgaa gaatgggttt 2340
gctgttgtga ggccccctgg ccatacagct gaagaatcca cagccatggg gttctgcttt 2400
tttaattcag ttgcaattac cgccaaatac ttgagagacc aactaaatat aagcaagata 2460
ttgattgtag atctggatgt tcacatgga aacggtaacc agcaggcctt ttatgctgac 2520
cccagcatcc tgtacatttc actccatcgc tatgatgaag ggaacttttt ccctggcagt 2580
ggagccccc aa atgaggttcg gtttatttct ttagagcccc acttttattt gtatctttca 2640
ggtaattgca ttgcatgatt acccctaatt ttcttgtcct ttgctgggtg tttaaattac 2700
acgagattac tgaattgtcc catgggacca agaaccagtg cagaacaagt gcataacca 2760
gagcactgtt tgcagggaa ggttgggctg atttgatgtg ttgtttgatg tttatttcaa 2820
gagctcccat gtgcttgttt tctctcttct ttgctttctt ccatttgctc tcttctctgc 2880
ccaccgtggt gtgtctttct cttcccaggt tggaaacaggc cttggagaag ggtacaatat 2940
aaatattgcc tggacaggtg gccttgatcc tcccatggga gatgttgagt acctgaagc 3000
attcaggacc atcgtgaagc ctgtggccaa agagtttgat ccagacatgg tcttagtate 3060
tgctggattt gatgcattgg aaggccacac ccctcctcta ggagggtaca aagtgcaggc 3120
aaaatgtttt ggtcatttga cgaagcaatt gatgacattg gctgatggac gtgtggtgtt 3180
ggctctagaa ggaggacatg atctcacagc catctgtgat gcacagaag cctgtgtaaa 3240
tgcccttcta ggaaatgagc tggagccact tgcagaagat attctccacc aaagcccgaa 3300
tatgaatgct gttatttctt tacagaagat cattgaaatt caaagtatgt ctttaaagtt 3360
ctcttaa

```

<210> 8  
 <211> 835  
 <212> PRT  
 <213> Homo sapiens

<400> 8  
 Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val  
 1 5 10 15  
 Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met  
 20 25 30  
 Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln  
 35 40 45  
 Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu  
 50 55 60  
 Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln  
 65 70 75 80  
 Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln  
 85 90 95  
 Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu  
 100 105 110  
 Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg  
 115 120 125

Gly 130	Lys 135	Asp 140	Arg 145	Gly 150	Arg 155	Glu 160	Arg 165	Ala 170	Val 175	Ala 180	Ser 185	Thr 190	Glu 195	Val 200	Lys 205
Gln 145	Lys 150	Leu 155	Gln 160	Glu 165	Phe 170	Leu 175	Leu 180	Ser 185	Lys 190	Ser 195	Ala 200	Thr 205	Lys 210	Asp 215	Thr 220
Pro 150	Thr 155	Asn 160	Gly 165	Lys 170	Asn 175	His 180	Ser 185	Val 190	Ser 195	Arg 200	His 205	Pro 210	Lys 215	Leu 220	Trp 225
Tyr 160	Thr 165	Ala 170	Ala 175	His 180	His 185	Thr 190	Ser 195	Leu 200	Asp 205	Gln 210	Ser 215	Ser 220	Pro 225	Pro 230	Leu 235
Ser 170	Gly 175	Thr 180	Ser 185	Pro 190	Ser 195	Tyr 200	Lys 205	Tyr 210	Thr 215	Leu 220	Pro 225	Gly 230	Ala 235	Gln 240	Asp 245
Ala 180	Lys 185	Asp 190	Asp 195	Phe 200	Pro 205	Leu 210	Arg 215	Lys 220	Thr 225	Glu 230	Ser 235	Val 240	Ser 245	Ser 250	
Ser 190	Ser 195	Pro 200	Gly 205	Ser 210	Gly 215	Pro 220	Ser 225	Ser 230	Pro 235	Asn 240	Asn 245	Gly 250	Pro 255	Thr 260	Gly 265
Ser 200	Val 205	Thr 210	Glu 215	Asn 220	Glu 225	Thr 230	Ser 235	Val 240	Leu 245	Pro 250	Pro 255	Thr 260	Pro 265	His 270	Ala 275
Glu 210	Gln 215	Met 220	Val 225	Ser 230	Gln 235	Gln 240	Arg 245	Ile 250	Leu 255	Ile 260	His 265	Glu 270	Asp 275	Ser 280	Met 285
Asn 220	Leu 225	Leu 230	Ser 235	Leu 240	Tyr 245	Thr 250	Ser 255	Pro 260	Ser 265	Leu 270	Pro 275	Asn 280	Ile 285	Thr 290	Leu 295
Gly 230	Leu 235	Pro 240	Ala 245	Val 250	Pro 255	Ser 260	Gln 265	Leu 270	Asn 275	Ala 280	Ser 285	Asn 290	Ser 295	Leu 300	Lys 305
Glu 240	Lys 245	Gln 250	Lys 255	Cys 260	Glu 265	Thr 270	Gln 275	Thr 280	Leu 285	Arg 290	Gln 295	Gly 300	Val 305	Pro 310	Leu 315
Pro 250	Gly 255	Gln 260	Tyr 265	Gly 270	Gly 275	Ser 280	Ile 285	Pro 290	Ala 295	Ser 300	Ser 305	Ser 310	His 315	Pro 320	His 325
Val 260	Thr 265	Leu 270	Glu 275	Gly 280	Lys 285	Pro 290	Pro 295	Asn 300	Ser 305	Ser 310	His 315	Gln 320	Ala 325	Leu 330	Leu 335
Gln 270	His 275	Leu 280	Leu 285	Leu 290	Lys 295	Glu 300	Gln 305	Met 310	Arg 315	Gln 320	Gln 325	Lys 330	Leu 335	Leu 340	Val 345
Ala 280	Gly 285	Val 290	Pro 295	Leu 300	His 305	His 310	Pro 315	Gln 320	Ser 325	Pro 330	Leu 335	Ala 340	Thr 345	Lys 350	Glu 355
Arg 290	Ile 295	Ser 300	Pro 305	Gly 310	Ile 315	Arg 320	Gly 325	Thr 330	His 335	Lys 340	Leu 345	Pro 350	Arg 355	His 360	Arg 365
Pro 300	Leu 305	Asn 310	Arg 315	Thr 320	Gln 325	Ser 330	Ala 335	Pro 340	Leu 345	Pro 350	Gln 355	Ser 360	Thr 365	Leu 370	Ala 375
Gln 310	Leu 315	Val 320	Ile 325	Gln 330	Gln 335	Gln 340	His 345	Gln 350	Gln 355	Phe 360	Leu 365	Glu 370	Lys 375	Gln 380	Lys 385
Gln 320	Tyr 325	Gln 330	Gln 335	Gln 340	Ile 345	His 350	Met 355	Asn 360	Lys 365	Leu 370	Leu 375	Ser 380	Lys 385	Ser 390	Ile 395
Glu 330	Gln 335	Leu 340	Lys 345	Gln 350	Pro 355	Gly 360	Ser 365	His 370	Leu 375	Glu 380	Glu 385	Ala 390	Glu 395	Glu 400	Glu 405
Leu 340	Gln 345	Gly 350	Asp 355	Gln 360	Ala 365	Met 370	Gln 375	Glu 380	Asp 385	Arg 390	Ala 395	Pro 400	Ser 405	Ser 410	Gly 415
Asn 350	Ser 355	Thr 360	Arg 365	Ser 370	Asp 375	Ser 380	Ser 385	Ala 390	Cys 395	Val 400	Asp 405	Asp 410	Thr 415	Leu 420	Gly 425
Gln 360	Val 365	Gly 370	Ala 375	Val 380	Lys 385	Val 390	Lys 395	Glu 400	Glu 405	Pro 410	Val 415	Asp 420	Ser 425	Asp 430	Glu 435
Asp 370	Ala 375	Gln 380	Ile 385	Gln 390	Glu 395	Met 400									

13/25

Gln Glu Thr Gly Leu Leu Asn Lys Cys Glu Arg Ile Gln Gly Arg Lys  
 625 630 635 640  
 Ala Ser Leu Glu Glu Ile Gln Leu Val His Ser Glu His His Ser Leu  
 645 650 655  
 Leu Tyr Gly Thr Asn Pro Leu Asp Gly Gln Lys Leu Asp Pro Arg Ile  
 660 665 670  
 Leu Leu Gly Asp Asp Ser Gln Lys Phe Phe Ser Ser Leu Pro Cys Gly  
 675 680 685  
 Gly Leu Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Leu His Ser Ser  
 690 695 700  
 Gly Ala Ala Arg Met Ala Val Gly Cys Val Ile Glu Leu Ala Ser Lys  
 705 710 715 720  
 Val Ala Ser Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro  
 725 730 735  
 Gly His His Ala Glu Glu Ser Thr Ala Met Gly Phe Cys Phe Phe Asn  
 740 745 750  
 Ser Val Ala Ile Thr Ala Lys Tyr Leu Arg Asp Gln Leu Asn Ile Ser  
 755 760 765  
 Lys Ile Leu Ile Val Asp Leu Asp Val His His Gly Asn Gly Thr Gln  
 770 775 780  
 Gln Ala Phe Tyr Ala Asp Pro Ser Ile Leu Tyr Ile Ser Leu His Arg  
 785 790 795 800  
 Tyr Asp Glu Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asn Glu Val  
 805 810 815  
 Arg Phe Ile Ser Leu Glu Pro His Phe Tyr Leu Tyr Leu Ser Gly Asn  
 820 825 830  
 Cys Ile Ala  
 835

&lt;210&gt; 9

&lt;211&gt; 1791

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 9

ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
 tgagggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggct 120  
 ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180  
 aagtcagaag ttctgtggg cctggagccc atctcacctt tagacctaa gacagacctc 240  
 aggatgatga tgcccggtgt ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
 cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
 cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaaactt 420  
 ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
 caagaacagg aagtagagag gcatcgagca gaacagcagc ttctctctct cagaggcaaa 540  
 gatagaggac gaaaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
 ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
 cgccatccca agctctggta cacggctgcc caccacacat cattggatca aagctctcca 720  
 ccccttagtg gaacatctcc atcttacaag tacacattac caggagcaca agatgcaaa 780  
 gatgatttcc cccttcgaaa aactgaatcc tcagtcagta gcagttctcc aggtctgtgt 840  
 cccagttcac caaacaatgg gccaaactgga agtgttactg aaaatgagac ttctggtttt 900  
 ccccctaccc ctcagtcgga gcaaatgggt tcacagcaac gcatttcta tcatgaagat 960  
 tccatgaacc tgctaagtct ttatacctct ccttctttgc ccaacattac cttgggggctt 1020  
 cccgcagtcg catccagct caatgcttgc aattcactca aagaaaagca gaagtgtgag 1080  
 acgcagacgc ttaggcaagg tgttctctg cctgggcagt atggaggcag catcccgga 1140  
 tcttcagacc accctcatgt tactttagag ggaaagccac ccaacagcag ccaccaggct 1200  
 ctccgtcagc atttattatt gaaagaacaa atgcgacagc aaaagcttct tgtagctggt 1260  
 ggagttccct tacatctca gtctcccttg gcaacaaaag agagaatttc acctggcatt 1320  
 agaggtaccc acaaatggcc ccgtcacaga cccctgaacc gaaccagtc tgcaccttg 1380  
 cctcagagca cgttggctca gctggctcatt caacagcaac accagcaatt cttggagaag 1440  
 cagaagcaat accagcagca gatccacatg aacaaactgc tttcgaaatc tattgaacaa 1500  
 ctgaagcaac caggcagtc ccttgaggaa gcagaggaa agcttcaggg ggaccaggcg 1560



14/25

```

atgcaggaag acagagcgcc ctctagtggc aacagcacta ggagcgacag cagtgccttgt 1620
gtggatgaca cactgggaca agttggggct gtgaagggtca aggaggaacc agtggacagt 1680
gatgaagatg ctcagatcca ggaaatggaa tctggggagc aggctgcttt tatgcaacag 1740
gtaataggca aagatttagc tccaggattt gtaattaaag tcattatctg a 1791

```

&lt;210&gt; 10

&lt;211&gt; 546

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

```

Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1      5      10      15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
20      25      30
Met Met Pro Val Val Asp Pro Val Val Arg Glu Lys Gln Leu Gln Gln
35      40      45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
50      55      60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65      70      75      80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
85      90      95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
100     105     110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
115     120     125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
130     135     140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145     150     155     160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
165     170     175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
180     185     190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
195     200     205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Glu Ser Ser Val Ser Ser
210     215     220
Ser Ser Pro Gly Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly
225     230     235     240
Ser Val Thr Glu Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala
245     250     255
Glu Gln Met Val Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met
260     265     270
Asn Leu Leu Ser Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu
275     280     285
Gly Leu Pro Ala Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys
290     295     300
Glu Lys Gln Lys Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu
305     310     315     320
Pro Gly Gln Tyr Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His
325     330     335
Val Thr Leu Glu Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu
340     345     350
Gln His Leu Leu Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val
355     360     365
Ala Gly Gly Val Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu
370     375     380
Arg Ile Ser Pro Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg
385     390     395     400
Pro Leu Asn Arg Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala

```

15/25

```

          405          410          415
Gln Leu Val Ile Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys
          420          425          430
Gln Tyr Gln Gln Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile
          435          440          445
Glu Gln Leu Lys Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu
          450          455          460
Leu Gln Gly Asp Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly
465          470          475          480
Asn Ser Thr Arg Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly
          485          490          495
Gln Val Gly Ala Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu
          500          505          510
Asp Ala Gln Ile Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met
          515          520          525
Gln Gln Val Ile Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val
          530          535          540
Ile Ile
545

```

<210> 11  
 <211> 590  
 <212> PRT  
 <213> Homo sapiens

```

<400> 11
Met His Ser Met Ile Ser Ser Val Asp Val Lys Ser Glu Val Pro Val
1          5          10          15
Gly Leu Glu Pro Ile Ser Pro Leu Asp Leu Arg Thr Asp Leu Arg Met
          20          25          30
Met Met Pro Val Val Asp Pro Val Arg Glu Lys Gln Leu Gln Gln
          35          40          45
Glu Leu Leu Leu Ile Gln Gln Gln Gln Ile Gln Lys Gln Leu Leu
          50          55          60
Ile Ala Glu Phe Gln Lys Gln His Glu Asn Leu Thr Arg Gln His Gln
65          70          75          80
Ala Gln Leu Gln Glu His Ile Lys Glu Leu Leu Ala Ile Lys Gln Gln
          85          90          95
Gln Glu Leu Leu Glu Lys Glu Gln Lys Leu Glu Gln Gln Arg Gln Glu
          100          105          110
Gln Glu Val Glu Arg His Arg Arg Glu Gln Gln Leu Pro Pro Leu Arg
          115          120          125
Gly Lys Asp Arg Gly Arg Glu Arg Ala Val Ala Ser Thr Glu Val Lys
          130          135          140
Gln Lys Leu Gln Glu Phe Leu Leu Ser Lys Ser Ala Thr Lys Asp Thr
145          150          155          160
Pro Thr Asn Gly Lys Asn His Ser Val Ser Arg His Pro Lys Leu Trp
          165          170          175
Tyr Thr Ala Ala His His Thr Ser Leu Asp Gln Ser Ser Pro Pro Leu
          180          185          190
Ser Gly Thr Ser Pro Ser Tyr Lys Tyr Thr Leu Pro Gly Ala Gln Asp
          195          200          205
Ala Lys Asp Asp Phe Pro Leu Arg Lys Thr Ala Ser Glu Pro Asn Leu
          210          215          220
Lys Val Arg Ser Arg Leu Lys Gln Lys Val Ala Glu Arg Arg Ser Ser
225          230          235          240
Pro Leu Leu Arg Arg Lys Asp Gly Asn Val Val Thr Ser Phe Lys Lys
          245          250          255
Arg Met Phe Glu Val Thr Glu Ser Ser Val Ser Ser Ser Ser Pro Gly
          260          265          270
Ser Gly Pro Ser Ser Pro Asn Asn Gly Pro Thr Gly Ser Val Thr Glu

```

16/25

275	280	285
Asn Glu Thr Ser Val Leu Pro Pro Thr Pro His Ala Glu Gln Met Val		
290	295	300
Ser Gln Gln Arg Ile Leu Ile His Glu Asp Ser Met Asn Leu Leu Ser		
305	310	315
Leu Tyr Thr Ser Pro Ser Leu Pro Asn Ile Thr Leu Gly Leu Pro Ala		
	325	330
Val Pro Ser Gln Leu Asn Ala Ser Asn Ser Leu Lys Glu Lys Gln Lys		
	340	345
Cys Glu Thr Gln Thr Leu Arg Gln Gly Val Pro Leu Pro Gly Gln Tyr		
	355	360
Gly Gly Ser Ile Pro Ala Ser Ser Ser His Pro His Val Thr Leu Glu		
	370	375
Gly Lys Pro Pro Asn Ser Ser His Gln Ala Leu Leu Gln His Leu Leu		
385	390	395
Leu Lys Glu Gln Met Arg Gln Gln Lys Leu Leu Val Ala Gly Gly Val		
	405	410
Pro Leu His Pro Gln Ser Pro Leu Ala Thr Lys Glu Arg Ile Ser Pro		
	420	425
Gly Ile Arg Gly Thr His Lys Leu Pro Arg His Arg Pro Leu Asn Arg		
	435	440
Thr Gln Ser Ala Pro Leu Pro Gln Ser Thr Leu Ala Gln Leu Val Ile		
	450	455
Gln Gln Gln His Gln Gln Phe Leu Glu Lys Gln Lys Gln Tyr Gln Gln		
465	470	475
Gln Ile His Met Asn Lys Leu Leu Ser Lys Ser Ile Glu Gln Leu Lys		
	485	490
Gln Pro Gly Ser His Leu Glu Glu Ala Glu Glu Glu Leu Gln Gly Asp		
	500	505
Gln Ala Met Gln Glu Asp Arg Ala Pro Ser Ser Gly Asn Ser Thr Arg		
	515	520
Ser Asp Ser Ser Ala Cys Val Asp Asp Thr Leu Gly Gln Val Gly Ala		
	530	535
Val Lys Val Lys Glu Glu Pro Val Asp Ser Asp Glu Asp Ala Gln Ile		
545	550	555
Gln Glu Met Glu Ser Gly Glu Gln Ala Ala Phe Met Gln Gln Val Ile		
	565	570
Gly Lys Asp Leu Ala Pro Gly Phe Val Ile Lys Val Ile Ile		
	580	585
		590

&lt;210&gt; 12

&lt;211&gt; 1084

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 12

Met Ser Ser Gln Ser His Pro Asp Gly Leu Ser Gly Arg Asp Gln Pro		
1	5	10
Val Glu Leu Leu Asn Pro Ala Arg Val Asn His Met Pro Ser Thr Val		
	20	25
Asp Val Ala Thr Ala Leu Pro Leu Gln Val Ala Pro Ser Ala Val Pro		
	35	40
Met Asp Leu Arg Leu Asp His Gln Phe Ser Leu Pro Val Ala Glu Pro		
	50	55
Ala Leu Arg Glu Gln Gln Leu Gln Gln Glu Leu Ala Leu Lys Gln		
65	70	75
Lys Gln Gln Ile Gln Arg Gln Ile Leu Ile Ala Glu Phe Gln Arg Gln		
	85	90
His Glu Gln Leu Ser Arg Gln His Glu Ala Gln Leu His Glu His Ile		
	100	105
Lys Gln Gln Gln Glu Met Leu Ala Met Lys His Gln Gln Glu Leu Leu		
		110



18/25

610	615	620
Arg Pro Leu Ser Arg Ala Gln Ser Ser Pro Ala Ser Ala Thr Phe Pro		
625	630	635
Val Ser Val Gln Glu Pro Pro Thr Lys Pro Arg Phe Thr Thr Gly Leu		
	645	650
Val Tyr Asp Thr Leu Met Leu Lys His Gln Cys Thr Cys Gly Ser Ser		
	660	665
Ser Ser His Pro Glu His Ala Gly Arg Ile Gln Ser Ile Trp Ser Arg		
	675	680
Leu Gln Glu Thr Gly Leu Arg Gly Lys Cys Glu Cys Ile Arg Gly Arg		
	690	695
Lys Ala Thr Leu Glu Glu Leu Gln Thr Val His Ser Glu Ala His Thr		
705	710	715
Leu Leu Tyr Gly Thr Asn Pro Leu Asn Arg Gln Lys Leu Asp Ser Lys		
	725	730
Lys Leu Leu Gly Ser Leu Ala Ser Val Phe Val Arg Leu Pro Cys Gly		
	740	745
Gly Val Gly Val Asp Ser Asp Thr Ile Trp Asn Glu Val His Ser Ala		
	755	760
Gly Ala Ala Arg Leu Ala Val Gly Cys Val Val Glu Leu Val Phe Lys		
	770	775
Val Ala Thr Gly Glu Leu Lys Asn Gly Phe Ala Val Val Arg Pro Pro		
785	790	795
Gly His His Ala Glu Glu Ser Thr Pro Met Gly Phe Cys Tyr Phe Asn		
	805	810
Ser Val Ala Val Ala Ala Lys Leu Leu Gln Gln Arg Leu Ser Val Ser		
	820	825
Lys Ile Leu Ile Val Asp Trp Asp Val His His Gly Asn Gly Thr Gln		
	835	840
Gln Ala Phe Tyr Ser Asp Pro Ser Val Leu Tyr Met Ser Leu His Arg		
	850	855
Tyr Asp Asp Gly Asn Phe Phe Pro Gly Ser Gly Ala Pro Asp Glu Val		
865	870	875
Gly Thr Gly Pro Gly Val Gly Phe Asn Val Asn Met Ala Phe Thr Gly		
	885	890
Gly Leu Asp Pro Pro Met Gly Asp Ala Glu Tyr Leu Ala Ala Phe Arg		
	900	905
Thr Val Val Met Pro Ile Ala Ser Glu Phe Ala Pro Asp Val Val Leu		
	915	920
Val Ser Ser Gly Phe Asp Ala Val Glu Gly His Pro Thr Pro Leu Gly		
	930	935
Gly Tyr Asn Leu Ser Ala Arg Cys Phe Gly Tyr Leu Thr Lys Gln Leu		
945	950	955
Met Gly Leu Ala Gly Arg Ile Val Leu Ala Leu Glu Gly Gly His		
	965	970
Asp Leu Thr Ala Ile Cys Asp Ala Ser Glu Ala Cys Val Ser Ala Leu		
	980	985
Leu Gly Asn Glu Leu Asp Pro Leu Pro Glu Lys Val Leu Gln Gln Arg		
	995	1000
Pro Asn Ala Asn Ala Val Arg Ser Met Glu Lys Val Met Glu Ile His		
	1010	1015
Ser Lys Tyr Trp Arg Cys Leu Gln Arg Thr Thr Ser Thr Ala Gly Arg		
1025	1030	1035
Ser Leu Ile Glu Ala Gln Thr Cys Glu Asn Glu Glu Ala Glu Thr Val		
	1045	1050
Thr Ala Met Ala Ser Leu Ser Val Gly Val Lys Pro Ala Glu Lys Arg		
	1060	1065
Pro Asp Glu Glu Pro Met Glu Glu Glu Pro Pro Leu		
	1075	1080

&lt;210&gt; 13

19/25

<211> 3550  
<212> DNA  
<213> Homo sapiens

<400> 13  
ggggaagaga ggcacagaca cagataggag aagggcaccg gctggagcca cttgcaggac 60  
tgaggggtttt tgcaacaaaa ccctagcagc ctgaagaact ctaagccaga tgggggtggc 120  
ggacgagagc agctcttggc tcagcaaaga atgcacagta tgatcagctc agtggatgtg 180  
aagtcagaag ttctgttggg cctggagccc atctcacctt tagacctaa gacagacctc 240  
aggatgatga tgcccgtggg ggaccctgtt gtccgtgaga agcaattgca gcaggaatta 300  
cttcttatcc agcagcagca acaaatccag aagcagcttc tgatagcaga gtttcagaaa 360  
cagcatgaga acttgacacg gcagcaccag gctcagcttc aggagcatat caaggaactt 420  
ctagccataa aacagcaaca agaactccta gaaaaggagc agaaactgga gcagcagagg 480  
caagaacagg aagtagagag gcatcgagca gaacagcagc ttctctctct cagaggcaaa 540  
gataaggagc gagaaagggc agtggcaagt acagaagtaa agcagaagct tcaagagttc 600  
ctactgagta aatcagcaac gaaagacact ccaactaatg gaaaaaatca ttccgtgagc 660  
cgccatccca agctctggta cagggtgccc caccacacat cattggatca aagctctcca 720  
ccccttagtg gaacatctcc atcctacaag tacacattac caggagcaca agatgcaaa 780  
gatgatttcc cccttcgaaa aactgcctct gagcccaact tgaaggtgag gtccagggtt 840  
aaacagaaag tggcagagag gagaagcagc cccttactca ggccggaagga tggaaatgtt 900  
gtcacttcat tcaagaagcg aatgtttgag gtgacagaat cctcagtcag tagcagttct 960  
ccaggctctg gtcccagttc accaaacaat gggccaactg gaagtgttac tgaaaatgag 1020  
acttcgggtt tggcccttac cctcatgccc gagcaaatgg ttccacagca acgcattcta 1080  
attcatgaag attccatgaa cctgctaagt ctttatacct ctcttctttt gcccaacatt 1140  
accttggggc ttcccgagc gccatcccag ctcaatgctt cgaattcact caaagaaaag 1200  
cagaagtgtg agacgcagac gcttaggcaa ggtgttcctc tgccctgggca gtatggaggc 1260  
agcatcccg catcttcag ccaccctcat gttacttttag agggaaagcc acccaacagc 1320  
agccaccagg ctctcctgca gcatttatta ttgaaagaac aaatgcgaca gcaaaagctt 1380  
cttgtagctg ttggagtcc cttacatcct cagtctccct tggcaacaaa agagagaatt 1440  
tcacctggca tagagggtac ccacaaattg ccccgtcaca gaccctgaa ccgaaccag 1500  
tctgcacctt tgcctcagag cacgttggct cagctggcca ttcaacagca acaccagcaa 1560  
ttcttgagga agcagaagca ataccagcag cagatccaca tgaacaaact gctttcgaaa 1620  
tctattgaac aactgaagca accaggcagt caccttgagg aagcagagga agagcttcag 1680  
ggggaccagg cgatgcagga agacagagcg ccctctagtg gcaacagcac taggagcgac 1740  
agcagtgcct gtgtggatga cacactggga caagtgggg ctgtgaaggt caaggaggaa 1800  
ccagtggaca gtgatgaaga tgctcagatc caggaaatgg aatctgggga gcaggctgct 1860  
tttatgcaac aggtaatagg caaagattta gctccaggat ttgtaattaa agtcattatc 1920  
tgacctttcc tggaaaccac gcacacacgt gcgctctctg tgcgccaagc tccgctggct 1980  
gcggttggca tggatggatt agagaaacac cgtctcgtct ccaggactca ctcttccct 2040  
gctgcctctg ttttacctca ccagcaatg gaccgcccc tccagcctgg ctctgcaact 2100  
ggaattgcct atgaccttct gatgtgaaa caccagtgcg tttgtggcaa ttccaccacc 2160  
caccctgagc atgctggacg aatacagagt atctggtcac gactgcaaga aactgggctg 2220  
ctaaataaat gtgagcgaat tcaaggtcga aaagccagcc tggaggaaat acagcttggt 2280  
cattctgaac atcactcact gttgtatggc accaaccctc tggacggaca gaagctggac 2340  
cccaggatac tcctaggtga tgactctcaa aagttttttt cctcattacc ttgtggtgga 2400  
cttgggggtg acagtgcac catttggaat gagctacact cgtccggtgc tgcacgcatg 2460  
gctgttggct gtgtcatcga gctggcttcc aaagtggcct caggagagct gaagaatggg 2520  
tttgctgttg tgaggcccc tggccatcac gctgaagaat ccacagccat ggggttctgc 2580  
ttttttaatt cagttgcaat taccgcaaaa tacttgagag accaactaaa tataagcaag 2640  
atattgattg tagatctgga tgttcaccat ggaaacggt cccagcaggc cttttatgct 2700  
gacccagca tcctgtacat ttactccat cgctatgatg aagggaactt tttccctggc 2760  
agtggagccc caaatgaggt tcggtttatt tcttttagag cccactttta tttgtatctt 2820  
tcaggtaatt gacttgcag attacccta attttctgt cctttgctgg tgttttaaat 2880  
tacacgagat tactgaattg tccatggga ccaagaacca gtgcagaaca agtgcataac 2940  
ccagagcact gtttgtcagg gaaggttggg ctgatttgat gtgttgtttg atgtttattt 3000  
caagagctcc catgtgcttg ttttctctc ttcttgcttt cttccatttg ctctcttctc 3060  
tgcccaccgt ggtgtgtctt tctcttccca ggttggaa ca ggccttgagg aagggtacaa 3120  
tataaatatt gcctggacag gtggccttga tctcccatg ggagatgttg agtaccttga 3180  
agcattcagg accatcgtga agcctgtggc caaagagttt gatccagaca tggctcttagt 3240  
atctgctgga tttgatgcat tgggaaggcca caccctcct ctaggagggt acaaagtgac 3300  
ggcaaaatgt tttggtcatt tgacgaagca attgatgaca ttggctgatg gacgtgtggt 3360  
gttggctcta gaaggaggac atgatctcac agccatctgt gatgcacag aagcctgtgt 3420

20/25

aatgcctt ctaggaaatg agctggagcc acttgcagaa gatattctcc accaaagccc 3480  
gaatatgaat gctgttattt ctttacagaa gatcattgaa attcaaagta tgtcttttaa 3540  
gttctcttaa 3550

&lt;210&gt; 14

&lt;211&gt; 7699

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 14

cccattcgcc attcaggctg cgcaactggt gggaagggcg atcgggtgagg gcctcttcgc 60  
tattacgccca gctggcgaaa ggggatgtg ctgcaaggcg attaagttgg gtaacgccc 120  
gggttttccc agtcacgacg ttgtaaaacg acggccagtg ccaagctgat ctaatcaata 180  
ttggccatta gccatattat tcattgggta tatagcataa atcaatattg gctattggcc 240  
attgcatacg ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt 300  
accgccatgt tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 360  
agttcatagc ccatatatgg agttccgcgt tacataactt acggtaaatg gccgcctgg 420  
cgaccgcccc gcgacccccg ccggttgacg tcaatagtga cgtatgttcc catagtaacg 480  
ccaataggga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg 540  
gcagtacatc aagtgtatca tatgccaaagt ccgcccccta ttgacgtcaa tgacggtaaa 600  
tgggccgcct agcattatgc ccagtacatg accttacggg agtttctac ttggcagtac 660  
atctacgtat tagtcatcgc tattaccatg gtgatgcggg ttggcagta caccaatggg 720  
cgtggatagc gggttgactc acgggggattt ccaagtctcc accccattga cgtcaatggg 780  
agtttgtttt ggcacaaaaa tcaacgggac ttccaaaaat gtcgtaataa ccccgccccg 840  
ttgacgcaaa tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta 900  
gtgaaccgtc agaattcaag cttgcggccg cagatctatc gatctgcagg atatcaccat 960  
gcacagtatg atcagctcag tggatgtgaa gtcagaagtt cctgtgggccc tggagcccat 1020  
ctcaccttta gacctaaagg cagacctcag gatgatgatg cccgtgggtgg accctgttgt 1080  
ccgtgagaag caattgcagc aggaattact tcttatccag cagcagcaac aaatccagaa 1140  
gcagcttctg atagcagagt ttcagaaaca gcagtgaagc ttgacacggc agcaccaggc 1200  
tcagcttcag gagcatatca aggaacttct agccataaaa cagcaacaag aactcctaga 1260  
aaaggagcag aaactggagc agcagaggca agaacaggaa gtagagaggc atcgagaga 1320  
acagcagctt cctcctctca gaggcaaaag tagaggacga gaaagggcag tggcaagtac 1380  
agaagtaaag cagaagcttc aagagttcct actgagtaaa tcagcaacga aagacactcc 1440  
aactaatgga aaaaatcatt ccgtgagccg ccattcccaag ctctgggtaca cggctgccc 1500  
ccacacatca ttggatcaaa gctctccacc ccttagtgga acatctccat cctacaagta 1560  
cacattacca ggagcacaag atgcaaagga tgatttcccc ctctgaaaaa ctgcctctga 1620  
gcccacttg aaggtgcggg ccagggttaa acagaaagtg gcagagagga gaagcagccc 1680  
cttactcagg cggaaggatg gaaatgttgt cacttcattc aagaagcgaa tgtttgaggt 1740  
gacagaatcc tcagtcatg gcagttctcc aggtcttggt ccagttcac caaacaatgg 1800  
gccaactgag agtggtactg aaaatgagac ttccggtttg cccctaccc ctcatgccga 1860  
gcaaatggtt tcacagcaac gcattctaat tcatgaagat tccatgaacc tgctaagttc 1920  
ttatacctct cttcttttgc ccaacattac ctctgggctt cccgcagtg catccagct 1980  
caatgcttcg aattcactca aagaaaagca gaagtgtgag acgcagacgc ttaggcaagg 2040  
tgttctctg cctgggcag atggaggcag catcccgga tctccagcc accctcatgt 2100  
tacttttagg ggaagccac ccaacagcag ccaccaggct ctctgcagc atttattatt 2160  
gaaagaacaa atgcgacagc aaaagcttct tgtagctggt ggagtccct tacatcctca 2220  
gtctcccttg gcaacaaaag agagaatttc acctggcatt agaggtaccc acaaattgcc 2280  
ccgtcacaga cccctgaacc gaacccagtc tgcaccttg cctcagagca cgttggctca 2340  
gctggtcatt caacagcaac accagcaatt cttggagaag cagaagcaat accagcagca 2400  
gatctcacatg aacaaactgc ttctgaaatc tattgaacaa ctgaagcaac caggcagtc 2460  
ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520  
ctctagtggc aacagcacta ggagcgacag cagtgcctgt gtggatgaca cactgggaca 2580  
agttggggct gtgaaggtca aggaggaacc agtgagacgt gatgaagatg ctcatatcca 2640  
ggaaatggaa tctggggagc aggtgtcttt tatgcaacag cctttccttg aaccacgca 2700  
cacacgtgcg ctctctgtgc gccaagctcc gctggctgcg gttggcatgg atggattaga 2760  
gaaacaccgt ctgctctcca ggactcactc tccccctgct gcctctgttt tacctcacc 2820  
agcaatggac cgccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880  
gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940  
acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000  
aggctgaaaa gccagcctgg aggaaatata gcttgttcat tctgaacatc actcactgtt 3060  
gtatggcacc aacccccctg acggacagaa gctggacccc aggatactcc taggtgatga 3120

ctctcaaaag tttttttcct cattaccttg tgggtggactt ggggtggaca gtgacaccat 3180  
ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240  
ggcttccaaa gtggcctcag gagagctgaa gaatgggttt gctgttgtag ggccccctgg 3300  
ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360  
cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420  
tcacatgga aacggtaccc agcaggcctt ttatgctgac ccagcatcc tgtacatttc 3480  
actccatcgc tatgatgaag ggaacttttt ccctggcagt ggagcccaa atgaggttgg 3540  
aacaggcctt ggagaaggg acaatataaa tatgcctgg acagggtggc ttgatcctcc 3600  
catgggagat gttgagtacc ttgaagcatt caggaccatc gtgaagcctg tggccaaaga 3660  
gtttgatcca gacatggtct tagtatctgc tggatttgat gcattggaag gccacacccc 3720  
tcctctagga ggtacaaaag tgacggcaaa atgttttggg catttgacga agcaattgat 3780  
gacattggct gatggacgtg tgggtgtggc tctagaagga ggacatgac tcacagccat 3840  
ctgtgatgca tcagaagcct gtgtaaatgc ccttctagga aatgagctgg agccacttgc 3900  
agaagatatt tcccaccaa gcccgaaatc gaatgctgtt atttctttac agaagatcat 3960  
tgaaattcaa agtatgtctt taaagtcttc tggatccggt accagattac aaggacgacg 4020  
atgacaagta gatcccgggt ggcacccctg tgaccctcc ccagtgcctc tcctggcctt 4080  
ggaagttgcc actccagtg cccaccagcct tgtcctaata aaattaagtt gcatcatttt 4140  
gtctgactag gtgtcctcta taatattatg ggtggaggg ggggtgtag gagcaagggg 4200  
cccaagttgg gaagacaacc tgtaggcct gcgggtcta ttcgggaacc aagctggagt 4260  
gcagtggcac aatcttggct cactgcaatc tccgcctcct gggttcaagc gattctcctg 4320  
cctcagcctc ccgagttgtt gggattccag gcattgcatga ccaggctcag ctaatttttg 4380  
tttttttggg agagacgggg tttcaccata ttggccaggc tggctccaa ctccataatc 4440  
caggtgatct acccaccttg gcctcccaa ttgctggat tacaggcgtg aaccactgct 4500  
ccctccctg tccttctgat tttaaaataa ctataccagc aggaggacgt ccagacacag 4560  
cataggctac ctgccatggc ccaaccggtg ggacatttga gttgcttgct tggcactgtc 4620  
ctctcatgcg ttgggtccac tcagtagatg cctgttgat tgggtacgag gccagcttct 4680  
gtggaatgtg tgtcagttag ggtgtgga gtcgccaggc tccccagcag gcagaagtat 4740  
gcaaagcatg catctcaatt agtcagcaac caggtgtgga aaagtccca ggctcccccag 4800  
caggcagaag tatgcaaac atgcatctca attagtacg aaccatagtc ccgcccctaa 4860  
ctccgcccac cccgccccta actccgcca gttccgcca ttctccgccc catggctgac 4920  
taattttttt tatttatgca gaggccgag cgcctcggc ctctgagcta ttccagaagt 4980  
agtgaggagg ctttttttga ggcctaggct tttgcaaaa gctcctcgag gaactgaaa 5040  
accagaaagt taattcccta tagtgagtcg tattaattc gtaatcatgg tcatagctgt 5100  
ttcctgtgtg aaattgttat ccgctcaca tttccacaa catacagacc ggaagcataa 5160  
agtgtaaagc ctgggtgcc taatgagtga gctaactcac attaatgagc ttgagctcac 5220  
tgcccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc ggccaacgag 5280  
cggggagagg cggtttgctg attgggcgct cttccgcttc ctgctcact gactcgctgc 5340  
gctcggctgc tcggctgcgg cgagcggat cagctcactc aaaggcggtg atacggttat 5400  
ccacagaatc aggggataac gcaggaaaaga acatgtgagc aaaaggccag caaaaggcca 5460  
ggaaccgtaa aaaggccgag ttgctggcgt ttttccatag gctccgcccc cctgacgagc 5520  
atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc 5580  
aggcgcttcc ccttggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg 5640  
gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta 5700  
ggatatctag ttcgggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg 5760  
ttcagcccg cccgtgcgct ttatccggtg actatcgtct tgagtccaac ccggtgaagc 5820  
acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag 5880  
gcggtgctac agagtctctg aagtgtgtgc ctaactacg ctacactaga agaacagtat 5940  
ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat 6000  
ccggcaaaac aaccaccgct ggtagcgggt gtttttttgt ttgcaagcag cagattacgc 6060  
gcagaaaaaa aggatctcaa gaagatcctt tgcatttttc tacggggtct gacgctcagt 6120  
ggaacgaaaa ctcacgttaa gggatttttg tcatgagatt atcaaaaagg atcttcacct 6180  
agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat gagtaaaact 6240  
ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctcagcgatc tgtctatttc 6300  
gttcatccat agttgcctga ctccccgtcg ttagataaac tacgatacgg gagggcctac 6360  
catctggccc cagtgtgca atgataccgc gagaccacg ctccaccggt ccagatttat 6420  
cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggctcctgca acttttatccg 6480  
cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg ccagttaata 6540  
gtttgogcaa cgttgttgcc attgctacag ccatcgttgt gtcacgctcg tcggttggtg 6600  
tggtttcatt cagctccggt tcccaacgat caaggcgagt tacatgatcc ccatgttgt 6660  
gcaaaaaagc ggttagctcc ttcggtcctc cgatcgttgt cagaagtaag ttggccgag 6720  
tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg ccatccgtaa 6780  
gatgcttttc tgtgactggg gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc 6840



22/25

gaccgagttg	ctcttgcccc	gcgtcaatac	gggataatac	cgcgccacat	agcagaactt	6900
taaaagtgtc	catcattgga	aaacgttctt	cggggcgaaa	actctcaagg	atcttaccgc	6960
tgttgagatc	cagttcgatg	taaccacatc	gtgcacccaa	ctgatcttca	gcattcttta	7020
ctttcaccag	cgtttctggg	tgagcaaaaa	caggaaggca	aaatgocgca	aaaaagggaa	7080
taagggcgac	acggaaatgt	tgaatactca	tactcttctt	ttttcaatat	tattgaagca	7140
tttatcaggg	ttattgtctc	atgagcggat	acataattga	atgtatttag	aaaaataaac	7200
aaataggggt	tccgcgcaca	tttccccgaa	aagtgccacc	tgacgcgccc	tgtagcggcg	7260
cattaagcgc	ggcgggtgtg	gtggttacgc	gcagcgtgac	cgctacactt	gccagcggcc	7320
tagcgcccg	tcctttcgct	ttcttccctt	cctttctcgc	cacgttcgcc	ggctttcccc	7380
gtcaagctct	aaatcggggc	atcccttttag	ggttccgatt	tagtgcttta	cggcacctcg	7440
accccaaaaa	acttgattag	ggtgatgggt	cacgtagtgg	gccatcgccc	tgatagacgg	7500
tttttcgccc	tttgacgttg	gagtcacagt	tctttaatag	tggactcttg	ttccaaactg	7560
gaacaacact	caaccctatc	tcggtctatt	cttttgattt	ataagggatt	ttgccgattt	7620
cggcctattg	gttaaaaaat	gagctgattt	aacaaaaatt	taacgcgaat	tttaacaaaa	7680
tattaaacgt	ttacaattt					7699

<210> 15  
<211> 7303  
<212> DNA  
<213> Homo sapiens

cccattcggc	attcaggctg	cgcaactggt	gggaaggggc	atcgggtgcg	gcctcttcgc	60
tattacgcca	gctggcgaaa	gggggatgtg	ctgcaaggcg	attaagttgg	gtaacgccc	120
gggttttccc	agtcacgacg	ttgtaaaacg	acggccagtg	ccaagctgat	ctaatacaata	180
ttggccatta	gccatattat	tcattgggtta	tatagcataa	atcaatattg	gctattggcc	240
attgcatacg	ttgtatccat	atcataatat	gtacatttat	attggctcat	gtccaacatt	300
accgccatgt	tgacattgat	tattgactag	ttattaatag	taatcaatta	cggggtcatt	360
agttcatagc	ccatataatg	agttccgcgt	tacataactt	acggtaaattg	gcccgcctgg	420
cgaccgcccc	gcgacccccg	cccggttgacg	tcaatagtga	cgtatgttcc	catagtaacg	480
ccaataggga	ctttccattg	acgtcaatgg	gtggagtatt	tacggtaaac	tgcccacttg	540
gcagtacatc	aagtgtatca	tatgccaagt	ccgcccccta	ttgacgtcaa	tgacggtaaa	600
tgccccgcct	agcattatgc	ccagtacatg	accttacggg	agtttctctac	ttggcagtac	660
atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggg	tttggcagta	caccaatggg	720
cgtggatagc	ggtttgactc	acggggattt	ccaagtctcc	acccatttga	cgtcaatggg	780
agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaa	gtcgtataaa	ccccgccccg	840
ttgacgcaaa	ttgggcggtag	gcgtgtacgg	ttggagggtc	atataagcag	agctcgttta	900
gtgaaccgtc	agaattcaag	cttgcggccg	cagatctatc	gatctgcagg	ataatccatt	960
gcacagtatg	atcagctcag	tggtatgtgaa	gtcagaagtt	cctgtggggc	tgagagccat	1020
ctcaccttta	gacctaagga	cagacctcag	gatgatgatg	cccgtggtgg	accctgttgt	1080
ccgtgagaag	caattgcagc	aggaattact	tcttatccag	cagcagcaac	aaatccagaa	1140
gcagcttctg	atagcagagt	ttcagaaaaca	gcattgagaa	ttgacacggc	agcaccaggc	1200
tcagcttcag	gagcatatca	aggaacttct	agccataaaa	cagcaacaag	aactcctaga	1260
aaaggagcag	aaactggagc	agcagaggca	agaacaggaa	gtagagaggc	atcgagagaa	1320
acagcagctt	cctcctctca	gaggcaaaga	tagaggacga	gaaagggcag	tggaaggtac	1380
agaagttaaag	cagaagcttc	aagagttcct	actgagtaaa	tcagcaacga	aagacactcc	1440
aactaatgga	aaaaatcatt	ccgtgagccg	ccatcccaag	ctctggtaca	cggctgcccc	1500
ccacacatca	ttggatcaaa	gctctccacc	ccttagtgga	acatctccat	cctacaagta	1560
cacattacca	ggagcacaa	atgcaaaagga	tgatttcccc	cttcgaaaaa	ctgcctctga	1620
gcccactttg	aaggtgcggg	ccagggttaa	acagaaagtg	gcagagagga	gaagcagccc	1680
cttactcagg	cgggaaggatg	gaaatgttgt	cacttcattc	aagaagcgaa	tgtttgaggt	1740
gacagaatcc	tcagtcagta	gcagttctcc	aggetctggt	cccagttcac	caaacaatgg	1800
gccaactgga	agtgttactg	aaaatgagac	ttcggttttg	ccccctacc	ctcatgccga	1860
gcaaatgggt	tcacagcaac	gcattctaat	tcattgaagat	tccatgaacc	tgctaagctc	1920
ttatacctct	ccttctttgc	ccaacattac	cttggggctt	cccgcagtcg	catcccagct	1980
caatgcttcg	aattcactca	aagaaaagca	gaagtgtgag	acgcagacgc	ttaggcaagg	2040
tgttctctcg	cctgggcagt	atggaggcag	catcccggca	tcttccagcc	accctcatgt	2100
tacttttagg	ggaaagccac	ccaacagcag	ccaccaggct	ctcctgcagc	attttattatt	2160
gaaagaacaa	atgcgacagc	aaaagcttct	tgtagctggt	ggagttccct	tacatcctca	2220
gtctcccttg	gcaacaaaag	agagaatttc	acctggcatt	agaggtaacc	acaaattgcc	2280

ccgtcacaga cccctgaacc gaaccagtc tgcacctttg cctcagagca cgttggctca 2340  
 gctgggtcatt caacagcaac accagcaatt cttgggagaag cagaagcaat accagcagca 2400  
 gatccacatg aacaaactgc tttcgaaatc tattgaacaa ctgaagcaac caggcagtc 2460  
 ccttgaggaa gcagaggaag agcttcaggg ggaccaggcg atgcaggaag acagagcgcc 2520  
 ctctagtggc aacagcacta ggagcgacag cagtgcctgt gtggatgaca cactgggaca 2580  
 agttggggct gtgaagggtca aggaggaacc agtggacagt gatgaagatg ctcagatcca 2640  
 ggaaatggaa tctggggagc aggtgctttt tatgcaacag cctttcctgg aaccacgca 2700  
 cacacgtgcg ctctctgtgc gccaaagctcc gctggctgcg gttggcatgg atggattaga 2760  
 gaaacaccgt ctctgtctcca ggactcactc ttccctgctt gcctctgttt tacctcacc 2820  
 agcaatggac cgccccctcc agcctggctc tgcaactgga attgcctatg accccttgat 2880  
 gctgaaacac cagtgcgttt gtggcaattc caccacccac cctgagcatg ctggacgaat 2940  
 acagagtatc tggtcacgac tgcaagaaac tgggctgcta aataaatgtg agcgaattca 3000  
 aggtcgaaaa gccagcctgg aggaataaca gcttgttcat tctgaacatc actcactgtt 3060  
 gtatggcacc aaccccctgg acggacagaa cctggacccc aggtactcc taggtgatga 3120  
 ctctcaaaaag tttttttcct cattaccttg tggtagactt ggggtggaca gtgacaccat 3180  
 ttggaatgag ctacactcgt ccggtgctgc acgcatggct gttggctgtg tcatcgagct 3240  
 ggcttccaaa gtggcctcag gagagctgaa gaattgggtt gctgttgtga ggccccctgg 3300  
 ccatcacgct gaagaatcca cagccatggg gttctgcttt tttaattcag ttgcaattac 3360  
 cgccaaatac ttgagagacc aactaaatat aagcaagata ttgattgtag atctggatgt 3420  
 tcaccatgga aacggtaccc agcaggcctt ttatgctgac ccagcatcc tgtacatttc 3480  
 actccatcgc tatgatgaag ggaactttt ccctggcagt ggagcccaa atgagggttcg 3540  
 gtttattttc ttagagccc acttttatt gtatctttca ggaattgca ttgcaggatc 3600  
 cggtaccaga ttacaaggac gacgatgaca agtagatccc ggtgggcatc cctgtgacc 3660  
 ctcccagtg cctctcctgg ccttggagt cgcactcca gtgcccacca gccttgcct 3720  
 aataaaaatta agttgcatca ttttgtctga ctagggtgcc tctataatat tatgggtgg 3780  
 aggggggtgg tatggagcaa ggggccaag ttgggaagac aacctgtagg gcctgcgggg 3840  
 tctattcggg aaccaagctg gagtgcagt gcacaatctt ggctcactgc aatctccgc 3900  
 tcctgggttc aagcgattct cctgcctcag cctcccgagt tgttgggatt ccaggcatgc 3960  
 atgaccaggc tcagctaatt tttgttttt ttgtagagac ggggtttcac catattggcc 4020  
 aggtctggct ccaactccta atctcagggt atctaccac cttggcctcc caaattgctg 4080  
 ggattacagg cgtgaaccac tgctccctc cctgtccttc tgattttaaa ataactatac 4140  
 cagcaggagg acgtccagac acagcatagg ctacttgcca tggcccaacc ggtgggacat 4200  
 ttgagttgct tgcttggcac tgcctctca tgcgttgggt ccactcagta gatgcctgtt 4260  
 gaattgggta cgcggccagc ttctgtggaa ttgtgtcag ttagggtgtg gaaagtcccc 4320  
 aggtcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag caaccagggtg 4380  
 tggaaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 4440  
 cagcaaccat agtcccgc ctaactccgc ccattcccgc cctaactccg cccagttccg 4500  
 cccattctcc gccccatggc tgactaattt tttttattta tgcagaggcc gaggccgct 4560  
 cggtctctga gctattccag aagtagtgag gaggttttt tggaggccta ggcttttgca 4620  
 aaaagctcct cgaggaactg aaaaaccaga aagttaattc cctatagtga gtcgtattaa 4680  
 attcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 4740  
 acaacatacg agccggaagc ataaagtgt aagcctgggg tgcctaata ga gtgagctaac 4800  
 tcacattaat tgcgttgccg tcaactgccg ctttccagtc gggaaacctg tctgtccagc 4860  
 tgcattaatg aatcggccaa cgcgcgggga gaggcggtt gcgtattggg cgtctctccg 4920  
 ctctcctcgt cactgactcg ctgcgctcgg tgcgtcggct gcggcgagcg gtatcagctc 4980  
 actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 5040  
 gagcaaaaag ccagcaaaaag gccaggaacc gtaaaaaggc cgcgttgcgt gcgtttttcc 5100  
 ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 5160  
 acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc gtgcgctctc 5220  
 ctgttccgac cctgcgcgtt accggatacc tgtccgcctt tctccctcgt ggaagcgtgg 5280  
 cgctttctca atgctcacgc ttaggtatc tcagttcggg taggtcgtt cgctccaagc 5340  
 tgggctgtgt gcacgaacct cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 5400  
 gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 5460  
 ggattagcag agcgagggtg gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 5520  
 acggttacac tagaagaaca gtatttggta tctgcgctct gctgaagcca gttaccttcg 5580  
 gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtgggtttt 5640  
 ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 5700  
 tttctacggg gtctgacgct cagtggacg aaaaactcag ttaagggatt ttggctatga 5760  
 gattatcaaa aaggatcttc acctagatcc ttttaaatga tttaaatcaa 5820  
 tctaaagtat atatgagtaa acttggctc acagttacca atgcttaatc agtgaggcac 5880  
 ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 5940  
 taactacgat acgggagggc ttaccatctg gccccagtc tgcaatgata ccgcgagacc 6000

24/25

```

cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 6060
gaagtgggcc tgcaacttta tccgcctcca tccagttctat taattgttgc cggaagccta 6120
gagtaagtag ttccgaggtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 6180
tggtgtcacg ctcgctgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 6240
gagttacatg atcccccatg ttgtgcaaaa aagcgggttag ctccctcggt cctccgatcg 6300
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 6360
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtag tcaaccaagt 6420
cattctgaga atagtgtatg cggcgaccga gttgtctctg cccggcgtca atacgggata 6480
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 6540
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaacct actcgtgcac 6600
ccaactgacg ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 6660
ggcaaaatgc cgcaaaaaag ggaataaggc cgacacggaa atgttgaata ctcatactct 6720
tcctttttca atattattga agcattttatc agggttattg tctcatgagc ggatacatat 6780
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 6840
cacctgacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt acgcgcagcg 6900
tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc ccttccttcc 6960
tcgccacggt cgccggcttt ccccgtaag ctctaaatcg gggcatccct ttagggttcc 7020
gatttagtgc tttacggcac ctcgacccca aaaaacttga ttaggggtgat ggttcacgta 7080
gtgggccatc gccctgatag acggtttttc gccctttgac gttggagtcc acgttcttta 7140
atagtggact cttgttccaa actggaacaa cactcaacct tatctcggtc tattcttttg 7200
atattataag gattttgccg atttcggcct attgggttaa aaatgagctg atttaacaaa 7260
aatttaacgc gaattttaac aaaatattaa acgtttacaa ttt 7303

```

&lt;210&gt; 16

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 16

ccatggaaac ggtacccagc aggc

24

&lt;210&gt; 17

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 17

cactccatcg ctatgatgaa ggg

23

&lt;210&gt; 18

&lt;211&gt; 23

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 18

agttcccttc atcatagcga tgg

23

&lt;210&gt; 19

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

25/25

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 19

aatgtacagg atgctggggt

20

&lt;210&gt; 20

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 20

cccttgtagc tggaggagtt ccctt

25

&lt;210&gt; 21

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 21

tgtgtcatcg agctggcttc

20

&lt;210&gt; 22

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Primer used to amplify human DNA

&lt;400&gt; 22

atcttctgca agtggctcca

20

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
27 December 2002 (27.12.2002)

PCT

(10) International Publication Number  
**WO 02/102984 A3**

- (51) International Patent Classification<sup>7</sup>: C12N 9/78, 9/00, 9/14, 1/20, 15/00, C07H 21/04
- (21) International Application Number: PCT/US02/19051
- (22) International Filing Date: 14 June 2002 (14.06.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/298,173 14 June 2001 (14.06.2001) US  
60/311,686 10 August 2001 (10.08.2001) US  
60/316,995 4 September 2001 (04.09.2001) US
- (71) Applicant (*for all designated States except US*): SLOAN-KETTERING INSTITUTE FOR CANCER RESEARCH [US/US]; 1275 York Avenue, New York, NY 10021 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): RICHON, Victoria [US/US]; 160 Theodore Fremd Street, #A11, Rye, NY 10580 (US). ZHOU, Xianbo [CN/US]; 43 Bradley Street, Dobbs Ferry, NY 10522 (US). RIFKIND, Richard, A. [US/US]; 425 East 58th Street, #48A, New York, NY 10022 (US). MARKS, Paul, A. [US/US]; 7 Rossiter Road, Washington, CT 06793 (US).
- (74) Agents: BROOK, David, E. et al.; Hamilton, Brook, Smith & Reynolds, P.C., 530 Virginia Road, P.O. Box 9133, Concord, MA 01742-9133 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— with international search report
- (88) Date of publication of the international search report:  
13 November 2003
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



WO 02/102984 A3

(54) Title: HDAC9 POLYPEPTIDES AND POLYNUCLEOTIDES AND USES THEREOF

(57) Abstract: The present invention features substantially pure HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), an HDRP( $\Delta$ NLS) polypeptides, and isolated nucleic acid molecules encoding those polypeptides. The present invention also features vectors containing HDAC9, HDAC9a, HDAC9( $\Delta$ NLS), HDAC9a( $\Delta$ NLS), and HDRP( $\Delta$ NLS) nucleic acid sequences, and cells containing those vectors.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/19054

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 9/78, 9/00, 9/14, 1/20, 15/00; C07H 21/04

US CL : 435/227, 183, 195, 252.3, 320.1; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/227, 183, 195, 252.3, 320.1; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
STN AND WEST. Sequence search in Swissprot, EST, N-GeneSeq, PIR\_71, SPTREMBL & issued US patents.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NAGASE et al. Prediction of Coding Sequences of Unidentified Human Genes. XI. The Complete Sequences of 100 New cDNA Clones from Brain Which Code for Large Proteins in Vitro. DNA Research November 1998, Vol 5, pages 277-286. See Table 1, Accession No. AB018287 is 58.8% similar to DNA sequence of SEQ ID NO : 1, claim 4 (g).	4
A, P	ZHOU et al. Cloning and Characterization of a histone deacetylase, HDAC9. PNAS, 11 September 2001, Vol. 98, No. 19, pages 10572-10577.	1-9, 29
A	WANG et al. HDAC4, a Human Histone Deacetylase Related to Yeast HDA1, Is a Transcriptional Corepressor. Molecular and Cellular Biology, November 1999, Vol. 19, No. 11, pages 7816-7827.	1-9, 29

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 October 2002 (30.10.2002)

Date of mailing of the international search report

13 MAR 2003

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Tekchand Saidha

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/190 51

## Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claim Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-9 & 29 (SEQ ID NOS : 1 & 2)

Remark on Protest

☐

The additional search fees were accompanied by the applicant's protest.

☐

No protest accompanied the payment of additional search fees.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/19051

### BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-9, 29, drawn to isolated nucleic acid, the encoded protein and protein composition.

Group II, claim(s) 10, drawn to antibody.

Group III, claim(s) 11-13, drawn to a method of identifying a compound - modulate DNA expression.

Group IV, claim(s) 14-19, 33, drawn to a method of identifying a compound that modulate enzymatic activity.

Group V, claim(s) 20-25, 34, drawn to a method of identifying a compound that modulate transcriptional repression activity of the polypeptide.

Group VI, claim(s) 26-27, drawn to a method of identifying a compound that modulate expression of a nucleic acid molecule.

Group VII, claim(s) 28, drawn to a method of identifying a polypeptide that interacts with a polypeptide of claim 1 in a two-hybrid system.

Group VIII, claim(s) 30-32, drawn to a method of diagnosing a cell proliferation disease.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack unity of invention because they are not so linked as to form a single general inventive concept under PCT Rule 13.1.

In order for more than one species to be examined, the appropriate additional examination fees must be paid. The species are as follows:

1. SEQ ID NO : 1 and 2 [HDAC9].
2. SEQ ID NO : 3 and 4 [HDAC9a].
3. SEQ ID NO : 5 and 6 [HDAC9- $\Delta$ NLS].
4. SEQ ID NO : 7 and 8 [HDAC9a- $\Delta$ NLS].
5. SEQ ID NO : 9 and 10 [HDRP- $\Delta$ NLS].

The claims are deemed to correspond to the species listed above in the following manner:

Each of the claims listed in groups I-VIII correspond to each of the 5 species which are structurally distinct.

The following claim(s) are generic: 1-5.

The inventions listed as Groups I-VIII do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Group I has a special technical feature of the nucleotide sequence encoding a specific histone deacetylase which Groups II-VIII do not share; Group II has a special technical feature of the antibody to a specific histone deacetylase which Groups I & III-VIII do not share; Groups III-VIII employ nucleic acid or polypeptide in various method of identifying compounds or polypeptides for distinct uses. Further, in view of 37 CFR 1.475 (b), when claims corresponding to different categories of inventions are present then only (3) applies and additional methods of use are deemed to lack unity.

The species listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: The various species correspond to nucleic acid and polypeptide sequences which are structurally and in activity distinct from each other, therefore lack the same or corresponding special technical feature.